

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 14:27:26 ; Search time 35.8 Seconds
(without alignments)
2935.079 Million cell

Title: US-09-840-746-1

perfect score: 4859

Sequence: 1 MSQTETVSRSVAPMRGGEIT.....FPGQYNPSFISDESRRRDYF 946

Scoring table:

PhyloSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Post processing: Minimum Match 0%
Maximum Match 100%

Maximum Match 100%
Listing first 45 summaries

Database :

1:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4798	98.7	1481	17	AAW03740	rchd528 gene product
2	4798	98.7	1481	20	AAW95160	Amino acid sequence
3	4798	98.7	1481	20	AAW93299	Human rchd528 gene
4	4798	98.7	1481	21	AAAB19626	Human rchd528 protein
5	4798	98.7	1481	21	AAAY78508	Amino acid sequence
6	4798	98.7	1481	21	AAAY68447	Rchd528 amino acid
7	1280	26.3	235	22	AAAG93453	Human polypeptide,
8	689	14.2	156	22	AAAB10235	Human CDNA SEQ ID
9	689	14.2	156	22	AAU18104	Novel human uterin
10	689	14.2	156	22	AAU17013	Human novel secret
11	689	14.2	156	22	AAU19898	Novel human calciu

ALIGNMENTS

RESULT

RESOLUTION
AAW03740
ID AAW03740 standard; Protein: 1481 AA.

AA
AC AAW03740:XX
DT 20-NOV-1996 (first entry)

XX DE rchd528 gene product.

XX Cardiovascular disease; differential expression; target gene;
KW pathway gene; fingerprint gene; atherosclerosis; ischaemia;
KW reperfusion; hypertension; restenosis; arterial inflammation;
KW vector; antibody; diagnosis; gene therapy; drug screening;
KW rchd528 gene.

XX
OS
Homo sapiens.

XX
FH
you
Location/Qualifiers

FH	key	Locat.
FT	Pentide	5 28

```
FT peptide
3...28
/label= sig peptide
```

FT	Region
/raser-319-peptide	1089..1122

```
FT
region
2000000000
/label= Extracellular_domain
```

FT /note="co

FT	re
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

FT	Modified-site	1140..1151
FT		(140) = 140

FT	/label= Asn_hydroxylation_site
FT	/note= "asparagine hydroxylation site consensus

FT /note="aspargine sequence"

FI	Domain	SE
ET	1348..1370	

```

11 domain
12 domain
13 domain
14 domain
15 domain
16 domain
17 domain
18 domain
19 domain
20 domain
21 domain
22 domain
23 domain
24 domain
25 domain
26 domain
27 domain
28 domain
29 domain
30 domain
31 domain
32 domain
33 domain
34 domain
35 domain
36 domain
37 domain
38 domain
39 domain
40 domain
41 domain
42 domain
43 domain
44 domain
45 domain
46 domain
47 domain
48 domain
49 domain
50 domain
51 domain
52 domain
53 domain
54 domain
55 domain
56 domain
57 domain
58 domain
59 domain
60 domain
61 domain
62 domain
63 domain
64 domain
65 domain
66 domain
67 domain
68 domain
69 domain
70 domain
71 domain
72 domain
73 domain
74 domain
75 domain
76 domain
77 domain
78 domain
79 domain
80 domain
81 domain
82 domain
83 domain
84 domain
85 domain
86 domain
87 domain
88 domain
89 domain
90 domain
91 domain
92 domain
93 domain
94 domain
95 domain
96 domain
97 domain
98 domain
99 domain
100 domain

```

PN WO9624604-A1.

PD 15-AUG-1996.
XX
PF 09-FEB-1996; 96WO-US01883.
XX
PR 07-JUN-1995; 95US-0485573.
PR 10-FEB-1995; 95US-0386844.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Falb DA;
XX
DR WPI; 1996-384391/38.
DR N-PSDB; AAT36035.
XX
PT New genes differentially expressed in cardiovascular disease - and
PT related vectors, host cells, proteins and antibodies, for diagnosis,
PT monitoring, treatment and drug screening
XX
PS Example 9; Fig 31A-D; 200pp; English.
XX
CC The amino acid sequence (AAW03740) was deduced of the protein encoded
CC by the novel human rchd528 gene (AAT36035). rchd528 is very highly
CC expressed in the heart. The rchd528 gene is up-regulated in
CC endothelial cells subjected to shear stress or exposed to oestrogen.
CC It is one of 8 novel genes (see also AAT36029-36) found to be
CC differentially expressed in cardiovascular disease (CVD), and is
CC useful in methods for the diagnosis and treatment of CVD. Assays
CC using cells expressing the novel genes can be used to identify cpds.
CC exhibiting anti-CVD activity.
XX
SQ Sequence 1481 AA;

Query Match 98.7%; Score 4798; DB 17; Length 1481;
Best Local Similarity 90.3%; Pred. No. 1.5e-257;
Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;

QY 1 MSQTETVSRVAPMRGEITAHLLNTSTTSADVTGSSSYPEGVNASVLTFQSDSTVQS 60
DB 436 msqtetvsvrsvapmrggeitahllntsttsadvtgssasypegvnasvltfqsdstvqs 495
QY 61 GCSHTALGDRSYSESSTSSSESLNSAPRGERS----- 94
DB 496 ggshtaigdrsyssesstssslnsaprgerstledsrepdgaigdsanaedrsgv 555
QY 95 ----- 94
DB 556 pslgthlatvtngertllrsvltnstsmsttsgeagspaaampqetegaslhvrvtdim 615
QY 95 -----IAGISGVQVRGTAIEQRTSSDHTDHTYLSSTFTKGERALLSITDNS 140
DB 616 glvrsiaaasalgvagisygqvrtaieqrtsdtdhtylsstftkgerallsitdns 675
QY 141 SSSDIVESSTSYIKINSSHSEYSSFSHAQTERSNTSSYDGEYAQPTESPVLHTSNLPS 200
DB 676 sssdivessstsyikinsnhseysfshaqtersnssydgeyaqptespvlhtsnlps 735
QY 201 YPTTINMPNTSVVLDTAERFVSDSSSSSSSSSGSGPPLPLPSVSQSHHLSFSSILPST 260
DB 736 ypttinpntsvvldtaerfvsssssssssgsgpplpplpsvsqshhlfssilpst 795
QY 261 RASVHLLKTSVSDATPWSSSPSPPLVSLTWTSTGAPLSVSGSTTLPQSSSTPVLPRARETPV 320
DB 796 rasvhllktsdastpwssspplvslttstgaptstaplsvsqsttlpqssstpvlpraretpv 855
QY 321 TSFQSTMTSFMPLHSSOPADILKSQSTPHQEKVITRESKSPSLVSLPTSTKAVTTNSPL 380
DB 856 tsfqsmtsfmplhssqcadlksqstphqekviteskspslvslptstkvttncspl 915
QY 381 PPSLTSSTSTQTLPATSTNIAQMSPTFTTTILKTSOPLMTTPGTLSTASLVTGPIAVQT 440
DB 916 ppsltsststqtlpatstnlqmsptftttlilksoplmttptgtlstaalvtgpiavqt 975

QY 441 TAGQLSLTHPEILVPOISTEGGISTERNRVIVDATTGLIPLTSVPTSAKEMTTKLGVTA 500
DB 976 tagqlslthpeilvpqisteggisternrvivdattgllpltsvptsakemttklgvt 1035
QY 501 EYSPASRSLGTSPTSPQTTVVSTAEADLPAPKATFAVQSGSTQSPPTLSSASVNSCAVPCL 560
DB 1036 eyspasrslgtspspqtvtvstaeadlpakstafavqsgstqpttllssasvnscavpcl 1095
QY 561 HNGECVADNTSRGYHCRCPSPSWQDDCSVDVNECLSNPCPSTATCNTTQGSFICKCPVG 620
DB 1096 hngecvadntsrghyrcrpspswqddcsvdvneclsnpcpstatcnttqgsfickcpvg 1155
QY 621 QLEKIGICNLVRTFVTEFKLRTFLNTTVEKHSDLQEVENEITKTLNMCFSALPSYIRSTV 680
DB 1156 qlekgicnlvrtfvtefkrlrtflnttvekhdsldqeveneitktnmcfsalpsyirstv 1215
QY 681 HASRESNAVVISLQTTFTSLASNVTLFDLADRMQKCVNSCKSSAEVCCOLLGSRQIFRAGS 740
DB 1216 hasresnavvislqttfslasnvtlfdladrmqkcvnsckssaevcqllgsgrrifrag 1275
QY 741 LCKRKSPEDKDTSICTDLGVALCCKSGVFOFNKMDHSCRACEDGYRLENETCMSCPF 800
DB 1276 lckrkspedkdtscitdldgvalccksgvfofnkmdhscracedgyrlenetcmscpf 1335
QY 801 GLGLNCGNPYQLITVYVIAAAGGGLLILGIALIVTCCRKNKNDISKLIKFGSDFQMSPY 860
DB 1336 glglnccnpyqlitvviaaaggllilgialivtccrknkndisklikfgsfqmspy 1395
QY 861 AEYPKNPRSQEWGREATEMHENGSTKNLLQMTDYYVYSPTSVRNPELERNGLYPAYTGLPG 920
DB 1396 aeypknprsqewgrealemhengstknllqmtdvyyspvsrnpelernglypaytglpg 1455
QY 921 SRHSCIFPGGYNPFSFISDESRRRDYF 946
DB 1456 srhscifpggynpfsfidesrrrdyf 1481

RESULT 2
AAW95160
ID AAW95160 standard; Protein; 1481 AA.
XX
AC AAW95160;
XX
DT 24-MAY-1999 (first entry)
XX
DE Amino acid sequence of rchd528 gene product.
XX
KW Fingerprinting gene; rchd502; transmembrane protein; cardiovascular;
KW fingerprint/target gene; up-regulated; endothelial cell; shear-stress;
KW atherosclerosis; ischemia; reperfusion; hypertension; restenosis.
XX
OS Homo sapiens.
XX
PN US5882925-A.
XX
PD 16-MAR-1999.
XX
PF 09-FEB-1996; 96US-0599654.
XX
PR 09-FEB-1996; 96US-0599654.
PR 10-FEB-1995; 95US-0386844.
PR 07-JUN-1995; 95US-0485573.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Falb DA;
XX
DR WPI; 1999-214071/18.
DR N-PSDB; AAX26246.
XX
PT New polynucleotides consisting of residues 1-1929 of the rchd502
PT gene - are differentially expressed in cardiovascular disease
PT states, and can therefore be used to treat and diagnose

PT cardiovascular diseases
 XX Disclosure; Fig 30A; 121pp; English.
 PS
 CC The invention relates to a rchd502 target/fingerprint gene encoding a
 CC transmembrane protein. The invention provides cDNAs contained in plasmids
 CC pFCHD502SF (ATCC 69981) and pFCHD502SJ (ATCC 69982) that encode the
 CC rchd502 polypeptide, and are differentially expressed in cardiovascular
 CC disease states. Cultured genetically engineered host cell containing the
 CC rchd502 polynucleotides in operative association with a nucleotide
 CC regulatory element are used for producing a polypeptide rchd502 gene
 CC product. Identifying that the fingerprint/target gene rchd502 is
 CC differentially expressed (up-regulated) by endothelial cells subjected
 CC to shear-stress, provides a tool for the diagnosis and treatment of
 CC cardiovascular disease e.g. atherosclerosis, ischemia/reperfusion,
 CC hypertension, restenosis. The fingerprint gene is useful for testing the
 CC efficacy of candidate drugs in basic research and in clinical trials and
 CC or imaging of a diseased cardiovascular tissue. The gene may also be
 CC used in screening for ligands of target gene product receptor domains, as
 CC well as antagonists of the ligand-receptor interaction.
 XX
 SQ Sequence 1481 AA;

Query Match 98.7%; Score 4798; DB 20; Length 1481;
 Best Local Similarity 90.3%; Pred. No. 1.5e-257;
 Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;

QY 1 MSQFTVSRVAPMRGGEITAHWLLTNTTSADVTGSSASYPEGVNASVLTFQSDSTVQS 60
 DB 436 msqtetvsvapmrgeitahwllntsttsadvtgssasypegvnasvltfqsdstvqs 495
 QY 61 GGSHTALGDRYSSESSSSSESLNSAPRGERS----- 94
 DB 496 ggshtaigdrysessstsseslnsaprgerstledrepqgalgdsanaedrtsgv 555
 QY 95 ----- 94

DB 556 pslgthlatvtgngertlrsvtlntsmsttsgeaaspaampqetegaslhvvnvtdm 615
 QY 95 -----IAGISYQVGRGTAIEQRTSSDHTDHTYLSSTFTKGERALLSITDNS 140
 DB 616 glvrslaaasalgvagisyygvrgtaieqrtssdhtdhtylsstftkgerallsidns 675
 QY 141 SSSDIVESSRYIKINSSHSEYFSFHAQTERSNISYDGEYAQPSTESPVLTHTSNLPS 200
 DB 676 sssdivessryiklnssshseyfsfhaqtersnissydgayqpstespvlhtsnlps 735
 QY 201 YTPITNPNTSVLDTDAEFVSDSSSSSSSSSSSGPPLPLPSVQSOSHHLPSSILPST 260
 DB 736 ytpitnmpntsvltdaefvsssssssssssgpplpplpsvqsqshlfsilpst 795
 QY 261 RASVHLKSTSDASTPSSSPPLVSLTSTTSAPLSVQTTLPOSSSTPVLPRARETPV 320
 DB 796 rasvhlkstdastpssspplvsltsttsaplsvsqtlpqssstpvlp raretpv 855
 QY 321 TSFOTSTMTSFWMLHSQRTADLKSQSTPQROEKVITESKSPSLVSLPSTKAVTNTNSPL 380
 DB 856 tsfotstmtsfmtlshsqrtadlksqstpqroekviteskpslvsplpstkavtntnsp 915
 QY 381 PPSLTESSTEQTLPATSTNLQAQMSPTTTTLTKTSQPLMTTPGTLSTASLVGTPIAVQT 440
 DB 916 ppsltessteqtlpatstnlqaqmspttttltktsqplmttptgtlssaslvgtpiavqt 975
 QY 441 TAGKQLSITHPEILVPOISTEGGISTERNRNVIVDATTGLIPLTSVPTSAKEMTTKLGVTA 500
 DB 976 tagqlslthpeilvpqisteggisternrvivdatglpltsvptsakemttklgvt 1035
 QY 501 EYSPASRLGTSPTQTVTVSTADLAPKATFAVQSTQSTPTTLSSASVNSCAVNPCL 560
 DB 1036 eyspasrlgtsptqtvvtvstadlapkatsfavqstqstpttlssasvnscavpnc 1095
 QY 561 HNGECVADNTSRGYHRCPPSWQGDGDCSDVNECLSNPCPSTATCNTQGSFICKCPVG 620

DB 1096 hngecvadntsrghrcppswqgdgdcsvdneclsnpcpstatcntqgsfickcpvg 1155
 QY 621 QLEKGICNLVTFVTEFKLRTFTLNTTVEKHSIDQEVENEITKTLMCMFSAALPSYIRSTV 680
 DB 1156 qlekgicnlvtfvtefkklrtfnttvekhsidqeveneitktlmmcfalsalpsyrstv 1215
 QY 681 HASRESNAVLSLOTTESLASNVTLFDLADRMQKCVNSCKSSAEVCCOLLGSSORIFRAGS 740
 DB 1216 hasresnavlsltttslasnvtfldladrmaqkcvnsckssaevcqlgsgrrifrag 1275
 QY 741 LCKRKSPECCKDTSICTDLDGVALCQCKSGYFQFNKMDHSCRACEDGYRLENETCMSCPF 800
 DB 1276 lckrkspecckdtsictldgvalcqcksgyfqfnkmdhscracedgyrlenetcmcpf 1335
 QY 801 GLGGLGCGNPYQLITVTAAGGGLLLILGIALIVTCCRNKNDISKLIKFGSGDFQMSPY 860
 DB 1336 glgglgcnpyqlitvvtaagggllilgialivtccrknkndisklikfgsgdfqmspy 1395
 QY 861 AEYKPNRQSQWEGREATEMHENGSTKNLLOMTDYYVYSPTSVRNPFLERNGLYPAYTGLPG 920
 DB 1396 aeykpnrsgwqgreatemhengstknllgmtdyyvyspvsrnpflernglypaytglpg 1455
 QY 921 SRHSCIFPGQYNPSFISDESRRDYF 946
 DB 1456 srhscifpgqynpsfidesrrdyf 1481

RESULT 3
 AAW89299
 ID AAW89299 standard; Protein; 1481 AA.
 XX
 AC AAW89299;
 XX
 DT 11-MAR-1999 (first entry)
 XX
 DE Human rchd528 protein.
 XX
 KW Human; cardiovascular disease; atherosclerosis; ischaemia; restenosis;
 KW reperfusion; hypertension; arterial inflammation; diagnosis; rchd528.
 XX
 OS Homo sapiens.
 XX
 PN US5849578-A.
 XX
 PD 15-DEC-1998.
 XX
 PF 15-MAR-1996; 96US-0616844.
 XX
 PR 09-FEB-1996; 96US-0599654.
 PR 10-FEB-1996; 95US-0386844.
 PR 07-JUN-1995; 95US-0458873.
 PR 15-MAR-1996; 96US-0616844.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Falb DA;
 XX
 XX WPI; 1999-069743/06.
 DR N-PSDB; AAV81825.
 DR
 XX
 PT DNA encoding rchd528 polypeptide - associated with cardiovascular
 PT disease
 XX
 XX Claim 1; Fig 30; 122pp; English.
 PS
 CC The present sequence represents rchd528 protein. A method has been
 CC developed for producing the rchd528 gene product. The present invention
 CC describes methods and compositions for the treatment and diagnosis of
 CC cardiovascular diseases, including: atherosclerosis; ischaemia;
 CC restenosis; reperfusion; hypertension; and arterial inflammation.
 XX
 SQ Sequence 1481 AA;

```

Query Match      98.7%; Score 4798; DB 20; Length 1481;
Best Local Similarity 90.3%; Pred. No. 1.5e-257;
Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;

QY 1 MSQTETVSRVAPMRGEITTAHLLTNTSTADVTGSSASYPEGVNASVLTFQFSDSTVQS 60
Db 436 msqetvsvravpmrgeitehwltnsttsadvtgssasypegvnasvltfqfstdstvs 495
QY 61 GGSHTALGDRSYSSSSSTSSSLSNSAPRGERS----- 94
Db 496 ggshtaigdrsyssssstsseslnssaprgerstledrepqgalgdsanaadrtsgv 555
QY 95 ----- 94
Db 556 pslgthlatvgtngertlrvltlntsmsttagaagspaampqetegaslhvntddm 615
QY 95 -----IAGISYQVGRGTAEQRTSSDHTDHTYLSSTFTKGERALLSIFDINS 140
Db 616 glvrsiaaasalgvagisygvrgtaeqrtssdhdtlylstftkgerallsicdms 675
QY 141 SSSDIVESSTYIKISNSHSEYSSFSHAQTERSNISSYDGEVAQSTESPVLHTSNLPS 200
Db 676 sssdivesscsyikinsnsheysfsfnagqcrsnissydgayaqstespvlhtsnlps 735
QY 201 YTPTINNPNTSVVLDTDAEFVSDSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSILPST 260
Db 736 ytptinmpntsvvltdaefvdsdssssssssssssssssssssssssssssssssllpst 795
QY 261 RASVHLKSTSDASTPWSSSPSPPLVSLTSTSAPLSVSQTTPLPQSSSTPVLPRARTPV 320
Db 796 rasvhlktsdstpwspspplvsltstsaplsvsttqplqssstpvlprrartpv 855
QY 321 TSFQTSMTSFMMLHSSQRTADLKSSQTPHOEKVITESKPSLSVSLPTESTKAVTTSNPL 380
Db 856 tsfqtstmsfmlhssqrtadlkssqtphoekviteskpslsvslpctskavttsnpl 915
QY 381 PPSLTESSTEQTLPATSTNLQAQSPFTTTLKTSQPLMTTPGTLSTSLASLTGPIAVQT 440
Db 916 ppsltesssteqtlpatstnlqaqspfttllktsqplmttptgtlslstasltvgpiavqt 975
QY 441 TACKQLSLTHPEILVPOISTEGGISTERNVIVDATTGLIPLISVPPTSAREMTKLGVTVA 500
Db 976 tagkqlslthpeilvpqisteggisternvividattgllpltsvptsaremtkklgvtva 1035
QY 501 EYSPASRLATSPSPQRTVYVSTADLAPKSGATFAVQSGSTQSPRTLSSASVNSCAVNPCL 560
Db 1036 eyspasrlatgspqrtvystadlapksgatfavqsgstqsprtlssasvnscavpncpl 1095
QY 561 HNGECVADNTSRGVCRCPPSWGGDCSDVDNECLSNPCPSTATCNNTQSGFICKCPVGY 620
Db 1096 hngecvadntsrghrcppswgddcsvdneclsnpcpstatcnntqsgfickcpvgy 1155
QY 621 QLEKGINLVRTFTVEFKLRTFLNTTVEKHSLOEVENETITLNMCFALSYIRSTV 680
Db 1156 qlekginlvrtftvefklrtflnttvekhsldqeveneitktlmcfsalpsyirstv 1215
QY 681 HASRESNAVVISLQTTTFLASNTVTLFOLADRMOKCVNSCKSSAEVCCOLLGQRRIFRAGS 740
Db 1216 hasresnavvislqtttflasnvtlfladrmqkcvnsckssaevcollgqrrifrag 1275
QY 741 LCKRKSPCEDKDTISICFDLGVLCQCKSGYFQFNKMDHSCRACEDGYRLNETCMSCPF 800
Db 1276 lckrkspcedkdtisicfdlglvalccksgyfqfnkmdhscracedyrlnetcmscpf 1335
QY 801 GLGLNCGNPYQLITVVVIAAGGELLILGIALIVTCCRNKNKNDISKLIFFKSGDFQMSPY 860
Db 1336 glglnccgncpyqlitvvvaaaaggllilgialivtccrknknndiskliffksgdfqmspy 1395
QY 861 AEPKPNRQSGEWGREATEMHENGSTKLLQMTDVIYSPTSVRNPFLERNGLIYATYGLPG 920
Db 1396 aepkpnrsqewgreaelemhengstknllqmtdvyyptsvrnpeflernglypaytglp 1455

```

```

QY 921 SRHSICFPQVNPSPFISDESRRRDYF 946
Db 1456 srhscifpgqynpsfidsesrrrdyf 1481

RESULT 4
AAB19626
ID AAB19626 standard; Protein; 1481 AA.
XX AAB19626;
XX 05-FEB-2001 (first entry)
XX Human rchd528 gene protein product.
XX Human; rchd528; differential expression; HUVEC; shear stress;
KW endothelial cell; cardiovascular disease; inflammation;
KW atherosclerosis; antiinflammatory; antiatherosclerotic; diagnosis;
KW therapy.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Peptide 5..28 /label= Signal_peptide
FT Protein 29..1431 /label= Mature_protein
FT Domain 1348..1370 /label= Transmembrane_domain
FT Region 1089..1122 /note= "epidermal growth factor repeat motif"
FT Modified-site 1140..1151 /note= "Asn is hydroxylated"
XX US6124433-A.
XX 26-SEP-2000.
XX 06-OCT-1997; 97US-0944496.
XX 09-FEB-1996; 96US-0599654.
XX 10-FEB-1995; 95US-0386844.
XX 07-JUN-1995; 95US-0485573.
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX (MILL-) MILLENNIUM PHARM INC.
XX Gimbrone MA, Falb DA;
XX WPI; 2000-611017/58.
XX N-PSDB; AAA88578.
XX Novel isolated rchd502 polypeptides, differentially expressed in
PT response to endothelial cell shear stress, used for diagnosis,
PT monitoring clinical trails, and treating cardiovascular diseases such
PT as ischemia -
XX Example 9.2; Fig 30A-I; 123pp; English.
XX The present sequence is that of the predicted polypeptide encoded
CC by the newly identified human rchd528 gene (see AAA88578). This gene
CC is differentially expressed (up-regulated) in response to endothelial
CC cell shear stress, and is not induced by interleukin-1. The
CC predicted protein is a novel G protein coupled receptor including 7
CC transmembrane domains. It shows 40% identity to the angiotensin II
CC receptor. Up-regulation of the rchd528 gene in a disease state may
CC reflect a protective role for the gene product in responding to
CC disease. Alternatively, it may have a causative or exacerbating
CC effect on the disease state. Modulation of rchd528 gene expression,
CC on the activity of its gene product, may provide a protective
CC effect. Knowledge of the gene and its protein product will provide
CC for drugs with greater specificity for the treatment of inflammation

```


CC and atherosclerosis. rchd528 is 1 of 8 novel human genes of the
CC invention (see AAA8576-83) characterized as being differentially
CC expressed in cardiovascular disease states, and which are of
CC diagnostic or therapeutic use.

XX	Sequence	1481 AA;	
SQ	Query Match	98.7%; Score 4798; DB 21; Length 1481;	
	Best Local Similarity	90.3%; Pred. No. 1.5e-257;	
	Matches	945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;	
QY	1	MSQTETVSRSVAPMRGGEITAHWLLTNTSTTSADVTSASYPEGVNASVLTQFSDSTVQS	60
DB	436	msqtetvsrvapmrgeitahwllntsttsadvtsasypegvnasvltqfstdstvsqs	495
QY	61	GGSHALGDRSYSSSSSTSSSESLSNSAPRGERS-----	94
DB	496	ggshtalgrdsysssstssseslnssaprgerstledsrepqgalgssanaedrtsgv	555
QY	95	-----	94
DB	556	pslghthlatvtngextlrsvltntsmsttsgeagspaaampqetegaslhvntddm	615
QY	95	-----IAGISYGOVRGTAIEQRTSSDHTDHTYLSSTFKGERALLSITDNS	140
DB	616	glvrsllaasalgvagisyyqvrgrtaeqdrtssdhcdhtyilstftkgerallscdns	675
QY	141	SSSDIVESSTYIKISNSHSEYSSFFSHAQTERSNISYDGEYAQPSTPVLHTSNLPS	200
DB	676	sssdivesstsyikinsnsheysffshaqternsissydgeyaqpstevplhtsnlps	735
QY	201	YTPINPNTSVLDTDAEFVSDSSSSSSSSSSSSSGPPLPLPVSQSHHLSFSSILPST	260
DB	736	ytptinmpntsvvldtdaefvdsdsssssssssssgpplplpsvqshhlfssilpst	795
QY	261	RASVHLKSTSDASTPWSSPPLPVSLTSTAPLSVSQTLTPQSSSTPVLPRARETPV	320
DB	796	rasvhlkstdastpwspsplpvsltstaplsvsqtltpqssstpvpraretpv	855
QY	321	TSFQTSMTSFMTHLSSQATADLKQSPTHQEVITESKPSLSVLPSTESTKAVTNSPL	380
DB	856	tsfqtstmtsfmthlssqatadlkqspthqeviteskpslsvlpstestkavtntsp1	915
QY	381	PPSLTESSTEQTLPATSTNLAQMSPTTTTLKTSQPLMTPTGTLSTASLVTCPIAVQT	440
DB	916	ppsltessteqtlpatstnlaqmspttttlktsqplmtptgtlslaslvtpgiavqt	975
QY	441	TAGKQLSLTHPEILVPOISTEGGISTERNRIVDATTGLIPLTSVPPTSAREMTTKLGVT	500
DB	976	tagkqlslthpeilvpisteggisternrivdattgllpltsvplsakemtcklgvta	1035
QY	501	EYSPASRLGTSPTQTVVSTVSTAEADLAPKATFAVQSTQSTPTLSSASVNSCAVNPCL	560
DB	1036	eyspasrlgtsptqtvvstvaeadlapkatsfavqstqstptllssasvnscavpnc1	1095
QY	561	HNGECVADNTPSRGYHRCPCPSWQDDCSVDVNECLSNPCSTATCNNTQSFICKCPVG	620
DB	1096	hngecvadtntsrghyrcpcpswqddcsvdvneclsnpcstacnntqsfickcpvg	1155
QY	621	QLKKGICNLVRTVFTEFKLRTFLNTTVKHSIDLQEVENETTKLNMCFSAIPSYIRSTV	680
DB	1156	qlekgicnlvrtvtefkrlrtflnttvkhsidlqeveneitklnmcfalspsyirstv	1215
QY	681	HASRESNAVVISLOTTFSLASNTVFLDLADRMQKVCVNSCKSSAEVCOLLGSQRIFRAGS	740
DB	1216	hasresnavvislqtfslasntvflldladrmlqvcvnsckssaevcqlgssqrifrag	1275
QY	741	LCRKRSPECCKDYSICTDLGVALCQCKSGYFQFNKMDHSCRACEDGYRLNETCMSCPF	800
DB	1276	lckrkspecckdysictdlgvalcqcksgyfgfnkmdhscracedgyrlenetcmscpf	1335
QY	801	GLGLGNCNPNQLITVIVAAAGGGLLLILGIALIVTCCRNKNDISKLIKSGDFQMSPY	860

DB	1336	glglgncnpyqlitvivaagggllilgialivtccrknkndiskliksgdfqmspy	1395
QY	861	AEYKPNRSPQEWGREAIEMHENGSTKNLQMTDYYYSPTSVRNPPELRNGLYPAYTGLPG	920
DB	1396	ayekpnrsqewgreaiemhengstknllqmtdvyysptsvrnpelernglypaytg1pg	1455
QY	921	SRHSCIFPGQYNPSFISDESRRRDYF	946
DB	1456	srhscifpgqynpsfisderrrdyf	1481
RESULT	5		
AA	AY78508		
ID	AA78508	standard; Protein; 1481 AA.	
XX	AC	AA78508;	
XX	AC		
DT	05-MAY-2000	(first entry)	
XX			
DE		Amino acid sequence of cardiovascular disease associated protein rchd528.	
XX			
KW		Differentially expressed; cardiovascular disease; atherosclerosis;	
KW		ischaemia; reperfusion; hypertension; restenosis; arterial inflammation;	
KW		rchd528; transmembrane protein.	
XX			
OS		Homo sapiens.	
XX			
PN		US6020463-A.	
XX			
PD		01-FEB-2000.	
XX			
PF		06-OCT-1997; 9705-0944423.	
XX			
PR		09-FEB-1996; 96US-0599654.	
PR		10-FEB-1995; 95US-0386844.	
PR		07-JUN-1995; 95US-0485573.	
XX			
PA		(BGHM) BRIGHAM & WOMENS HOSPITAL.	
PA		(MILL-) MILLENNIUM PHARM INC.	
XX			
PI		Gimbrone MA, Falb DA;	
DR		WPI: 2000-146911/13.	
DR		N-PSDB; AA289796.	
XX			
PT		Marker proteins for the diagnosis of cardiovascular diseases such as	
PT		atherosclerosis and hypertension, comprising peptide sequences derived	
PT		from the rchd523 transmembrane protein -	
XX			
PS		Examples; Fig 20; 121pp; English.	
XX			
CC		This sequence represents the human rchd528 amino acid sequence. This	
CC		sequence is related to the rchd523 transmembrane polypeptide which is	
CC		encoded by cDNA contained in the plasmid pfchd523. The rchd523 protein is	
CC		differentially expressed in diseased cells compared to healthy cells. The	
CC		rchd523 protein may be used as a marker protein for the diagnosis of	
CC		cardiovascular diseases including atherosclerosis, ischaemia,	
CC		reperfusion, hypertension, restenosis and arterial inflammation. rchd523	
CC		peptides may be used as antigens in the production of antibodies specific	
CC		for rchd523. The anti-rchd523 antibodies may then be used in diagnostic	
CC		assays to quantitate rchd523 peptides in samples.	
XX			
SQ		Sequence 1481 AA;	

Query Match	98.7%;	Score 4798;	DB 21;	Length 1481;
Best Local Similarity	90.3%;	Pred. No. 1.5e-257;		
Matches	945;	Conservative 1;	Mismatches 0;	Indels 100; Gaps 1;
QY	1	MSQTETVSRSVAPMRGGEITAHWLLTNTSTTSADVTSASYPEGVNASVLTQFSDSTVQS	60	
DB	436	msqtetvsrvapmrgeitahwllntsttsadvtsasypegvnasvltqfstdstvsqs	495	

Qy	61	GGSHtALCDRbSYSESStSSSESLNSsAPRGERS-----	94
Db	496	ggshTALgDRsySSStSSSESLNSsAPRGerStledsreppgqalGdsanaedrTsgv	555
Qy	95	-----	94
Db	556	pslgthclatvtgngertlrsvtlntsmsttsgeagspaaampqetegaslhvnrvtddm	615
Qy	95	-----IAGISyGOVRGTAIEQRtSSDHtTYLSStFKGERALLSIFDNS	140
Db	616	glvrsGLaaasalgvaglyggvrgtaieqrtsdhtdhtylstfckgerallsidcns	675
Qy	141	SSSDIVESStSYIKINSsHSySFSHAQrTsrNISsYdGEYAQpStESPVLtSNLPS	200
Db	676	ssadivEsstsykklnsnshesyfsfhaqrtnsnissydgeyqdpstespvlhtsnlps	735
Qy	201	YtPTINmPntsvVLDPAEFvSDSSSSSSSSSSSGGpPLPsvSQSHHLSFSLPESt	260
Db	736	YtptInmPntsvvldcaefvdsdsssssssssgsggpllpvsgshlfsellpSt	795
Qy	261	RASvHLLKStSOASTPWSsPpLVsVLTtTSPALPSYsOttLpQSSStPVLPRARETPV	320
Db	796	rasvhllKstsdastpwspspplvstltsaplvsvqtlpqssstvpwlraretpv	855
Qy	321	TSFQtStMTsFMtMLHSSoTABLKsQStPHOEKVITEskSPSLVsLStESTKAVtTNSPL	380
Db	856	tsfqtstmtsfmLmlhsqtadlksgstphqekviteskpslvslptestkavttnspl	915
Qy	381	PPSltESSStEOTLPATstNLAOmStFTtTlKtSQPLMTtPGtLStTASlVtGPiAVQT	440
Db	916	ppsltessceqtLpactslagmsptfttlktsqplmttpgtltstasltvtpgiavqt	975
Qy	441	TAGKQLSLThPEllVpQISTEGIGstERNRrVtDAtGLIPLTSVpTSAKEMtTKLGvTA	500
Db	976	tagkqlslthpeellvpqistegigsternrvldatgtlptsvptsakemtktlgvtA	1035
Qy	501	EYSPASrSLGtSPSPottVVtAEDLAKsATFAVOSStQSPtTLSSASvNSCAVNpCL	560
Db	1036	eyspasrSLgtspspqtvtvstaedlApksatfvgsqtgpttlssasvnsCAVnPCl	1095
Qy	561	HNCEvADNtSRGyHCRCPpSQGDGCSVDVNECLNSPCPtATCNtQGSfICKCPVGy	620
Db	1096	hngecvADntsrgyhrcpPpSvqgdcsvDvneclnsnpCpstaCnntqgsfICKcpvgY	1155
Qy	621	QLEBKICNLVrtVtFEfKLKRtFLNtTVeKHSDlQEVENEItKTLNMCfSALPSYIRStV	680
Db	1156	qlEkglcnlvrtvtfefkkrftflnttvekhshdlqeveneItkTLnmcfaalpsyirStv	1215
Qy	681	HASRESNAVVISLQTFtSLASNVtLFDLADRMKVCNSKSSAEVcOLLGSORIFRAGS	740
Db	1216	hasresNAVvislqtctfsLasnvtlfdlAdrmqkvcnckssaeVcqlLsgqrfrfrags	1275
Qy	741	LCRRKSPEDKDtSICtDLdGVALCQCKSGYFQFNKMbHSCRACEDGYrLENETCMScPF	800
Db	1276	lckrkspEdkDtslctdldgvalcqcksgyfQfnkmhdsCracedgyrLenetcmScpf	1335
Qy	801	GLGGLNCGNPYQLITVvIAAGGGLLLlIGtIALIVtCCrKNKNDISKLIFKSGDFQMSPY	860
Db	1336	glgglncGnpYqlitvviaaeggllllglIALivtccrknkndisklIFKsgdfqmspy	1395
Qy	861	AEYPKPNRSQEGREAIEMHENGStKNLQMTDvYtSPtSVrNPELERNGLYPAYtCLPG	920
Db	1396	aeypkPNrsqEGreaiemhengstknllqMTdvYtSPtSVrnpelernGLypaytCLpg	1455
Qy	921	SRHSCIFPGQYNPStfSDSERRDYF	946
Db	1456	srhscIfpgqYnpStfSDSerrdyf	1481

RESULT 6

RESUL
AAY68447

AA10047
ID AAY68447 standard: protein: 1481 AA.

XX	AAAY68447;	
AC		
XX		
DT	19-APR-2000	(first entry)
XX		
DE	Rchd528	amino acid sequence SEQ ID NO:40.
XX		
KW	Cardiovascular disease;	diagnosis; atherosclerosis; ischaemia;
KW	refusion; hypertension;	stenosis; arterial inflammation;
KW	antiartherosclerotic;	vasotropic; hypotensive.
XX		
OS	Homio sapiens.	
XX		
PN	US6018025-A.	
XX		
PD	25-JAN-2000.	
XX		
PF	06-OCT-1997;	97US-0944868.
XX		
PR	09-FEB-1996;	96US-0599654.
PR	10-FEB-1995;	95US-0386844.
PR	07-JUN-1995;	95US-0485573.
XX		
PA	(MILL-) MILLENUM PHARM INC.	
PA	(BGHM) BRIGHAM & WOMENS HOSPITAL.	
XX		
PI	Falb DA,	Gimbrone MA;
XX		
DR	WPI;	2000-136704/12.
DR	N-PSDB;	AAZ88007.
XX		
PT	Isolated polypeptide	for treating and diagnosing cardiovascular
PT	disease, such as,	atherosclerosis, ischemia/reperfusion, hypertension,
PT	restenosis and	arterial inflammation -
XX		
PS	Claim 1;	Fig 30; 122pp; English.
XX		
CC	The present invention	describes an isolated polypeptide (I) comprising
CC	either the amino acid	sequence of 1481 residues, given in AAY68447, or
CC	an amino acid sequence	encoded by the cDNA contained in plasmids
CC	pFCHD528A (ATCC 63985),	pFCHD528B (ATCC 63986) and pFCHD528C (ATCC
CC	63987). The polypeptide	is useful in the treatment and diagnosis of
CC	cardiovascular disease,	such as, atherosclerosis, ischaemia/reperfusion,
CC	hypertension, restenosis	and arterial inflammation. AAZ88001 to AAZ88040,
CC	and AAY68444 to	AAY68457 represent sequences used in the exemplification
CC	of the present	invention.
XX		
SQ	Sequence	1481 AA;
	Query Match	98.7%; Score 4798; DB 21; Length 1481;
	Best Local Similarity	90.3%; Pred. NO. 1.5e-257;
	Matches 945;	Conservative 1; Mismatches 0; Indels 100; Gaps 1
QY	1	MSQTEVRSVAPMRGGEITAHLLTNSTTSADVTGSSASYPEGVNASVLTQFSDSVQS 60
Db	436	msqtetvrsvapmrgeitahlltnsttsadvtgssasypegvnaavltqfstdstvs 495
QY	61	GGSHALGRDSYSESSSTSSSELNSSAPRGERS----- 94
Db	496	ggshalgrdsysssstssselnsasprgerstledrepgqalgdsanaedrsgv 555
QY	95	----- 94
Db	556	pslghthlatvtgngertlrsvtlntsmsttsgeagspaaapqpetegashlvnvtddm 615
QY	95	-----IAGISYGOVGRGTAEQRTSSDHTDHTYLSSTFKGERALLSIDNS 140
Db	616	glvrslaassalgvagisvgvrgcaterqtssdhtcdhtylssftkgerallsidns 675
QY	141	SSSDIVESSTSYIKINSNSHSYISFSHAQTERSNISSYDGEYIAQPTSEPLVLTSLNLS 200
Db	676	sssdivesstsviklansshsvssfhagtersnissvddvgaqstcesvplvtslnls 735

XX	02-AUG-2001.		PR	29-SEP-2000;	2000US-236367P.
PD			PR	29-SEP-2000;	2000US-236368P.
XX			PR	29-SEP-2000;	2000US-236369P.
XX	17-JAN-2001;	2001WO-US01349.	PR	29-SEP-2000;	2000US-236370P.
XX			PR	02-OCT-2000;	2000US-236802P.
PR	31-JAN-2000;	2000US-179065P.	PR	02-OCT-2000;	2000US-237037P.
PR	04-FEB-2000;	2000US-180628P.	PR	02-OCT-2000;	2000US-237038P.
PR	24-FEB-2000;	2000US-184664P.	PR	02-OCT-2000;	2000US-237039P.
PR	02-MAR-2000;	2000US-186350P.	PR	13-OCT-2000;	2000US-237040P.
PR	16-MAR-2000;	2000US-189874P.	PR	13-OCT-2000;	2000US-239935P.
PR	17-MAR-2000;	2000US-190076P.	PR	13-OCT-2000;	2000US-239937P.
PR	18-APR-2000;	2000US-198112P.	PR	20-OCT-2000;	2000US-240960P.
PR	19-MAY-2000;	2000US-205513P.	PR	20-OCT-2000;	2000US-241221P.
PR	07-JUN-2000;	2000US-209467P.	PR	20-OCT-2000;	2000US-241785P.
PR	28-JUN-2000;	2000US-214886P.	PR	20-OCT-2000;	2000US-241786P.
PR	30-JUN-2000;	2000US-215135P.	PR	20-OCT-2000;	2000US-241787P.
PR	07-JUL-2000;	2000US-216647P.	PR	20-OCT-2000;	2000US-241808P.
PR	07-JUL-2000;	2000US-216880P.	PR	20-OCT-2000;	2000US-241809P.
PR	11-JUL-2000;	2000US-217487P.	PR	20-OCT-2000;	2000US-241826P.
PR	11-JUL-2000;	2000US-217496P.	PR	01-NOV-2000;	2000US-244617P.
PR	14-JUL-2000;	2000US-218290P.	PR	08-NOV-2000;	2000US-246474P.
PR	26-JUL-2000;	2000US-220963P.	PR	08-NOV-2000;	2000US-246475P.
PR	26-JUL-2000;	2000US-220964P.	PR	08-NOV-2000;	2000US-246476P.
PR	14-AUG-2000;	2000US-224518P.	PR	08-NOV-2000;	2000US-246477P.
PR	14-AUG-2000;	2000US-224519P.	PR	08-NOV-2000;	2000US-246478P.
PR	14-AUG-2000;	2000US-225213P.	PR	08-NOV-2000;	2000US-246523P.
PR	14-AUG-2000;	2000US-225214P.	PR	08-NOV-2000;	2000US-246524P.
PR	14-AUG-2000;	2000US-225266P.	PR	08-NOV-2000;	2000US-246525P.
PR	14-AUG-2000;	2000US-225267P.	PR	08-NOV-2000;	2000US-246526P.
PR	14-AUG-2000;	2000US-225288P.	PR	08-NOV-2000;	2000US-246527P.
PR	14-AUG-2000;	2000US-225270P.	PR	08-NOV-2000;	2000US-246528P.
PR	14-AUG-2000;	2000US-225447P.	PR	08-NOV-2000;	2000US-246532P.
PR	14-AUG-2000;	2000US-225757P.	PR	08-NOV-2000;	2000US-246609P.
PR	14-AUG-2000;	2000US-225758P.	PR	08-NOV-2000;	2000US-246610P.
PR	14-AUG-2000;	2000US-225759P.	PR	08-NOV-2000;	2000US-246611P.
PR	28-AUG-2000;	2000US-226279P.	PR	08-NOV-2000;	2000US-246613P.
PR	22-AUG-2000;	2000US-226681P.	PR	17-NOV-2000;	2000US-249207P.
PR	22-AUG-2000;	2000US-226868P.	PR	17-NOV-2000;	2000US-249208P.
PR	22-AUG-2000;	2000US-227182P.	PR	17-NOV-2000;	2000US-249209P.
PR	30-AUG-2000;	2000US-228924P.	PR	17-NOV-2000;	2000US-249210P.
PR	01-SEP-2000;	2000US-229287P.	PR	17-NOV-2000;	2000US-249211P.
PR	01-SEP-2000;	2000US-229343P.	PR	17-NOV-2000;	2000US-249212P.
PR	01-SEP-2000;	2000US-229344P.	PR	17-NOV-2000;	2000US-249213P.
PR	01-SEP-2000;	2000US-229345P.	PR	17-NOV-2000;	2000US-249214P.
PR	05-SEP-2000;	2000US-229509P.	PR	17-NOV-2000;	2000US-249215P.
PR	05-SEP-2000;	2000US-229513P.	PR	17-NOV-2000;	2000US-249216P.
PR	06-SEP-2000;	2000US-230437P.	PR	17-NOV-2000;	2000US-249217P.
PR	06-SEP-2000;	2000US-230438P.	PR	17-NOV-2000;	2000US-249218P.
PR	08-SEP-2000;	2000US-231242P.	PR	17-NOV-2000;	2000US-249244P.
PR	08-SEP-2000;	2000US-231243P.	PR	17-NOV-2000;	2000US-249245P.
PR	08-SEP-2000;	2000US-231244P.	PR	17-NOV-2000;	2000US-249264P.
PR	08-SEP-2000;	2000US-231413P.	PR	17-NOV-2000;	2000US-

DR WPI: 2001-476161/51.
DR N-PSDB; ABA06457.
XX
PT . Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
XX
PS Claim 11; SEQ ID NO: 543; 859pp + Sequence Listing; English.
XX
XX
CC The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.
XX
XX
SQ Sequence 156 AA;

Query Match 14.2%; Score 689; DB 22; Length 156;
Best Local Similarity 85.0%; Pred. No. 5.2e-31;
Matches 130; Conservative 3; Mismatches 16; Indels 4; Gaps 1;

QY 496 LGVTAEYSPASRLGTSPTQTTWSTAEDLAPKSAFVQSSQTQSTPTLSSASVNSCA 555
I : |||||
Db 4 laiqstspasrlgtsptqttwstaedlapksatfavqstqstptlssasvnsca 63
QY 556 VNPCLHNGECVADNTRGYYHCRPPSWQGDGCSVDVNECLSNPCPSATCNNTQGSFICK 615
I : |||||
Db 64 vnpclhngecvadtngyhcrcppswqgdcsvdvneclsnpcpsatcnntqgxxick 123
QY 616 CPVGYQLKGCICNLVRFVTFEFLKRTFLNTTV 648
I : |||||
Db 124 cpvgyqlkxlcnlk----rlxlftrftrti 152

RESULT 9
AAU18104
ID AAU18104 standard; Protein; 156 AA.
XX
AC AAU18104;
XX
XX
DT 21-NOV-2001 (first entry)
XX
XX
DE Novel human uterine motility-association polypeptide #11.
XX
XX
KW Human; uterine motility-association disorder; uterus; pregnancy;
KW labour; menstrual cycle; gene therapy.
XX
OS Homo sapiens.
XX
XX
PN WO200155201-A1.
XX
XX
PD 02-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-0501317.
XX
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180828.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225269.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.

PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0234274.
PR 14-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0237935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476222/51.
XX N-PSDB; AAS26918.

Novel polypeptides and polynucleotides useful as diagnostic reagents to diagnose diseases or disorders associated with aberrant expression or activity of polypeptides, for treating blood clotting disorder, haemophilia

Claim 11; SEQ ID No 254; 601pp; English.

The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used

CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
Query Match 14.2%; Score 689; DB 22; Length 156;
Best Local Similarity 85.0%; Pred. No. 5.2e-31;
Matches 130; Conservative 3; Mismatches 16; Indels 4; Gaps 1;
QY 496 LGVTAEVSPASRLGTPSPQTTVSTAEADLPKSAATEAVQSSSTQSTPTTLSSASVNSCA 555
Db 4 laiqqstspasrlgtpspqttvstaeadlpkscatfavqscqstpttlssasvnsca 63
QY 556 VNPCLHNGECVADNTRGYHCRPPSQWQGDSCVDVNECLNCPSTATCNCNQTGSGFICK 615
Db 64 vnpclhngecvadtargyhcrppsqwqgdscvdvneclnncpstatcncnqtgsgxxick 123
QY 616 CPVGYQLKGCINLVTFVTEFLKRTFLNTTV 648
Db 124 cpvgqylekxicnlgk-----rlxlftrltfrtti 152
RESULT 11
AAU19898
ID AAU19898 standard; Protein; 156 AA.
XX AC AAU19898;
XX DT 06-DEC-2001 (first entry)
XX DE Novel human calcium-binding protein #7.
XX KW Human: calcium-binding protein; calcium flux; neurological disease;
KW Immune dysfunction; digestive disorder; neoplastic disease;
KW blood disorder; infectious disease; gene therapy; immunosuppressive;
KW antiarthritic; cytostatic; vasotropic; antibacterial; nootropic;
KW virucide.
XX OS Homo sapiens.
XX WO200155304-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01302.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465568/50.
XX N-PSDB; AAS31583.

Isolated nucleic acid molecule encoding a calcium-binding protein is used in preventing, treating or ameliorating a medical condition -

Claim 11; SEQ ID No 95; 542pp; English.

XX The present invention relates to the isolation of novel human
XX calcium-binding proteins, and cDNA (AAS31577-AAS31654) and genomic
XX sequences encoding for these proteins. The sequences of the invention
XX are useful in the diagnosis, prevention and/or prognosis of diseases
XX associated with aberrant calcium flux. Such disorders include
XX neurological diseases (e.g. amyotrophic lateral sclerosis, ALS),
XX immune dysfunction (e.g. severe combined immunodeficiency, SCID),
XX digestive disorders (e.g. irritable bowel syndrome, IBS), neoplastic
XX disease (e.g. cancer), blood disorders (e.g. haemophilia), and/or
XX infectious disease (e.g. acquired immunodeficiency syndrome, AIDS). The
XX novel calcium-binding proteins are also useful as screening tools to
XX identify antagonists and/or agonists that may enhance or inhibit
XX activities mediated by calcium-binding proteins. The polynucleotides of
XX the invention are also useful in gene therapy. AAU19892-AAU19969
XX represent the novel human calcium-binding proteins.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 156 AA;

Query Match 14.2%; Score 689; DB 22; Length 156;
Best Local Similarity 85.0%; Pred. No. 5.2e-31;
Matches 130; Conservative 3; Mismatches 16; Indels 4; Gaps 1;

Qy 496 LGVTAEYSPASRLGTSPTQTVVSTAEELAPKATFAVQSSTQSPPTLSSASVNSCA 555
Db 4 laiqgstspasrlgtspsqtvtvstaeelapksatfavqsstqspptlssasvnsca 63
Qy 556 VNPCLHNGECVADNTRSGYHCRPCPPSQWQDCSDVNECLSNPCPSTATCNNTCQSFICK 615
Db 64 vnpcLhngECvADntrSGYhCRpCpPsQwQdCsDvNEclSNpCpStAtCnnTcQsFiCk 123
Qy 616 CPVGYQLEKIGICNLVTRTFVTEFKLRKRTFLNTTV 648
Db 124 cpvgYqLeKixicnlgk----rlxlftrtfrtti 152

RESULT 12

AAU16941
ID AAU16941 standard; Protein; 121 AA.

XX
AC AAU16941;

XX
DT 07-NOV-2001 (first entry)

XX
DE Human novel secreted protein, SEQ ID 182.

XX Human; immunosuppressive; antiarthritic; antirheumatic;
KW cystostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.

XX OS Homo sapiens.

XX PN WO200155441-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01320.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 14-AUG-2000; 2000US-0225759.

XX PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239355.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-476222/51.
 DR N-PSDB; AAS26846.
 XX
 XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
 PT diagnose diseases or disorders associated with aberrant expression or
 PT activity of polypeptides, for treating blood clotting disorder,
 PT haemophilia
 XX
 PS Claim 11; SEQ ID No 182; 601pp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC

Query Match 10.8%; Score 523; DB 22; Length 121;
 Best Local Similarity 89.8%; Pred. No. 5.9e-22;

Matches 97; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 510 CTSPPQTIVVSTADLAPKATRAVAGSTQSPPTLSSASVNCVAVNCPCLHNGECVADN 569
 Db 1 gtsppqgtvvstaeatlakfatfavgstgspptlssasvncvavncpclhngcecvadn 60

QY 570 TSGVYHCRCPSPGWDGDCSDVDNCLSNPCPSTATCNCNTQGSFICKCP 617
 Db 61 tsrgyhcrppswgddcdvnciclnpcbstatwgqysgillylqmp 108

RESULT 13
 AAU37120
 ID AAU37120 standard; Protein; 2344 AA.
 AC AAU37120;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Staphylococcus aureus cellular proliferation protein #1290.
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Staphylococcus aureus.
 XX
 PN W0200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 WP: 2001-611495/70.
 DR N-PSDB; AAS54979.
 XX
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 12713; 51lpp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, themselves and the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX

SQ Sequence 2344 AA;

Query Match 7.8%; Score 377; DB 22; Length 2344;
 Best Local Similarity 24.0%; Pred. No. 3.8e-12;
 Matches 195; Conservative 149; Mismatches 350; Indels 120; Gaps 24;

QY 2 SQTETVRSVAPMRGGEITAHWLLTNSTTSADV-----TCSSASYPGVNASVLTQFS 54
 Db 1054 sdsdkslsldsgsgts-----tstssvrmsesqsgsmstsgdsstsf 1108

QY 55 DSTVQGGSHALGDRSYSESSSTSSSESLNS---SAPRCERSIAGISYGVQVRCFAIEQ 110
 Db 1109 dstdsksasta-sseislsqsvstsgsvstslstlscnertstmsdtslstsseed 1167

QY 111 RTSSDHTDHTYLSSTFTKGERALLSITDNSSSDI-VESSTSYIKIS---NSSHSEYSF 166
 Db 1168 stdststsdiseaisgsestsislscnstsdsesksasafleslsestsestseal 1227

QY 167 SHAOTERSNISSYDGEYAQPSTE-----SPVLHTSNLPS-YPTINMPNTSVVLDTD- 217
 Db 1228 sgstsdtsldnsesgtsstlsnstsasistsgsaststvkseavstslstst 1287

QY 218 AEFVSDSSSSSSSSSSSGPPLPL-PSVQSHHLFSSILPSTRASVHLLKSTSDA--- 273
 Db 1288 stslsdtsldtsldsgsgsknsleamstsdsktrkesisastslsgstseseg 1347

QY 274 STPWSSSPS---PLPVSLTTSAPLSVSQT-TLPQSSSTPVLPRAR-----EPVTSF 323
 Db 1348 stssesksdtsmlsmgsgtsvstseslstdtsstslsasnmqgvdnsasq 1407

QY 324 QTSMTSPMTMLHSSQ-TADLKQSQTPHOEKVITESKPSLSVLPTESTKAVTINSPLPP 382
 Db 1408 sastststsdsgstsytsqsdtsstslsdtsksksgststas 1467

QY 383 SLTSESTEQTLPATSTNLQMSPTFTTILKTSOPLMTTPTGLSTASLVTPGPIAVQTTA 442
 Db 1468 sgseesdsgstsesksestslsdts-----tsnsgsaatstllnsaasesd 1523

QY 443 GKQLSLTHPEILVPQISTEGGISTERNRIVDATTGLIPLTSVPT-SAKEMTTKLGVTAE 501
 Db 1524 ssslsldstsaamqsgsesdsgststslnsqgststslrmstlasesvsestseegstse 1583

QY 502 YSPASRSLGTSPS-PQTTVVSTFAEDLAPKSAFVQSQSTQSPPTLSSASV----- 551
 Db 1584 stsesdtsldsdsgstsrsts---asgsaststsdtsrstaastststldsgs 1640

QY 552 -----NSCAVNPCLHNGECVADNTSRGYHCRCPSPSWGDDCSDVDNCLSNPCP-STAT 604
 Db 1641 mslststsvdtsldsvdts-----dststsgsmaasistldst 1687

QY 605 CNNTQGSFICKPCVGYQLEKICNLVRTFTVFTEFKLRTFLNTVEKHSDLOEVEWEITKT 664
 Db 1688 ststasevmsasi-----sdsqsmseavnds 1714

QY 665 LNMCFPSALPYIRTVTHASRESNAVVISLQTF-----SLASNVTILFDLADRMQKVCNSCK 720
 Db 1715 evsesnesdskmsdsvsgstsvslrksesvsesislsqsgsmndsvstsd 1774

QY 721 SSAEVCQLLGSRRIFRAGSLCKRKSPECDKDT 754
 Db 1775 sslsvstsqrsseavsesdlsdskstsgststs 1808

RESULT 14
 AAU59288
 ID AAU59288 standard; Protein; 957 AA.
 XX
 AC AAU59288;
 XX
 DT 25-APR-2000 (first entry)
 XX
 DE Human MUC11 polypeptide.

XX Mucin; MUC11; MUC12; human; chromosome 7q22; epithelial inflammation;
 KW Crohn's disease; ulcerative colitis; asthma; chronic bronchitis;
 KW colorectal cancer; cystic fibrosis; inflammatory bowel disease;
 KW breast cancer.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH Modified-site 68
 FT /note: "potential N-glycosylation site"
 FT
 PN W0200004142-A1.
 XX
 XX 27-JAN-2000.
 XX
 XX 16-JUL-1999; 99WO-AU00579.
 XX
 XX 16-JUL-1998; 98AU-0004708.
 XX
 XX (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.
 PA (ORDE-) ORDER OF SISTERS OF MERCY IN QUEENSLAND.
 XX
 XX Williams SJ, Antalis TM, McGuckin MA, Gotley DC;
 XX WPI; 2000-182416/16.
 XX N-PSDB; AA258824.
 XX
 XX Novel MUC nucleic acid corresponding to mucin gene, useful for treating
 PT associated disease conditions e.g. colorectal, breast cancer, cystic
 PT fibrosis and inflammatory bowel disease -
 XX
 XX Claim 10; Page 83-88; 103pp; English.
 XX
 CC The invention provides mucin genes (MUC11 and MUC12) located on human
 CC chromosome 7q22. The mucin genes or its portion is used in detecting
 CC polymorphism, mutation, deletion, truncation and expansion in the gene
 CC or its gene transcript. Pharmaceutical compositions and gene therapy
 CC constructs comprising the mucin genes are used for treating disease
 CC conditions associated with aberrant Mucin expression, altered properties
 CC of mucus or epithelial inflammatory processes involving Mucins like
 CC Crohn's disease, ulcerative colitis, asthma, chronic bronchitis and
 CC colorectal cancer, cystic fibrosis, inflammatory bowel disease and breast
 CC cancer. The mucin genes and the polypeptides are used for determining
 CC these diseases or their predisposition. The MUC11 and MUC12 polypeptides
 CC are used for preparing antagonist and antibodies. The present sequence
 CC represents the human MUC11 polypeptide.
 XX
 XX Sequence 957 AA;
 SQ
 Query Match 6.7%; Score 326; DB 21; Length 957;
 Best Local Similarity 26.1%; Pred. No. 7.8e-10;
 Matches 168; Conservative 84; Mismatches 243; Indels 148; Gaps 28;
 QY 28 STTSADVTGSSASYPGVNALSIVLQFSDSTVQSG-----GSHTPALG--DRSYSESSSTS 79
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 342 sttsgvseestshspgthtctafpdtsttptglshsttshspgtdtlltpasttt 401
 QY 80 ---SSSLNSSAPRGERSIAGISGVQVRGAIEORTSSDHTD----HTYLSSTTKGERA 132
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 402 sgpsgeatshspgtdta-lspgstalsfgqesttshspgsthltl----- 450
 QY 133 LLSLTDNSSSDIYESSTYKISNHSSEYSSPSHA-----QTERSNISSYDGE-YA 184
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 451 ---fpdsttsgliveast---rvhsstgsprttlspasstspglqgestafqtphastht 504
 QY 185 QPST-----ESPVLHTSNLPSYPTTNMPTSVV-----LDTDAEFVSDSSSSSS 230
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 505 tpsstpatapveesttyhrs--psstptthfpasssttshgskstifhspdsagttps 562
 QY 231 SSSSSSGPPLPLPSVQSQHLHFFSILPSTRASVHLKST-----SDASPWSSSP-SPL 284
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 6. . .

DB 563 sahattag-----rgesttarispgstteittlpgstttlpglseastfyssprsp 613
 QY 285 ----PVSL-----TTSTSAPLSVSQTLTPOSSSSPVLPRARETPTVTFQSTTM-- 329
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 614 tllspasmtslgvgeesttsrqgsthstvspasttllpglseesttysspsgstetv 673
 QY 330 ---SFTMTLHSSQADLKSQSTPHQEKVITESKPSLSLPTST-----KAVTNSPLPP 382
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 674 fprstttsvrggepttthfshpashthtltfstdstts--glteestafpgspastqtgipa 731
 QY 383 SLT-----SSSTEQTL-PATSTNL-----AQMSPPTFTTILKTSQPLM 419
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 732 tlttadlgeesttffsssgstgttllsparsttsglvgestsrisspsstettllpgsp-- 789
 QY 420 TTPGTLSTASIVTGPITAVQTTAGKQLSLTHPEILLVPOISTEGGISTERNRIVVDATGL 479
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 790 tplselseksttffysprspdat-----lspatttssgvsee-----sstshs 831
 QY 480 IPLTSVPTSAKEMTTKLGVTAEYSPASRSLGTSPSPQTVTVSTARDLAPKSAFAVQSST 539
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 832 qpsthtttafpdtsttsglsqepkthshsqgsteatlspgattasslqgqsttfh-sapg 890
 QY 540 QSPPTLSSASVNSCAVNP-----CLHNGECVADNTSRG 573
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 891 dtettllpddtitsglveastpshstsgslhtltltpasstsaag 933
 RESULT 15
 AAM24513
 ID AAM24513 standard; Protein; 957 AA.
 XX
 AC AAM24513;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE C900P predicted amino acid sequence.
 XX
 KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 KW immunogenic; gene therapy; vaccine; colonic cancer.
 XX
 OS Homo sapiens.
 XX
 PN W0200149716-A2.
 XX
 PD 12-JUL-2001.
 XX
 PF 29-DEC-2000; 2000WO-US35596.
 XX
 PR 30-DEC-1999; 99US-0476296.
 PR 10-JAN-2000; 2000US-0480321.
 PR 15-FEB-2000; 2000US-0504629.
 PR 06-MAR-2000; 2000US-0519444.
 PR 19-MAY-2000; 2000US-0575251.
 PR 29-JUN-2000; 2000US-0609448.
 PR 28-AUG-2000; 2000US-0649811.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
 PI King GE, Wang T, Jiang Y;
 XX
 XX WPI; 2001-441847/47.
 XX
 PT Colon tumor associated proteins and nucleic acids useful for the
 PT prevention, diagnosis and treatment of colonic cancer -
 XX
 PS Claim 2; Page 437-440; 472pp; English.
 XX
 CC The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)

Search completed: July 23, 2002, 14:30:06
Job time: 160 sec

```
Query Match      6.7%; Score 326; DB 22; Length 957;
Best Local Similarity 26.1%; Pred. No. 7.Be-10;
Matches 168; Conservative 84; Mismatches 243; Indels 148; Gaps 28;
```

Qy	28	STTSADVTGSSASYPEGVNASVLTFQSDSVQSG-----GSHTALG--DRSYSESSSTS	79
Db		: : : : : : :	
Qy	342	stssgvseestshsrpggsthtafdpdsttgpqlsthssthspsgstdtlipasttt	401
Db		: : : : : : :	
Qy	80	--SSELNSAPRGERSIAGISYGOVRGTAEIQRTSSDHTD----HTYLSSTFTKGERA	132
Db		: : : : : : : : :	
Qy	402	sgpsqestahspgsdtda-lpggstataifggesttfhspgsthttl-----	450
Db		: : : : : : : : :	
Qy	133	LLSTDNSSDDIVESSTSYTKINSRSHVESVSPSHA-----OTERNSTSSVDGE-YA	184
Db		: : : : : : :	
Qy	451	--fpdstasglveast---rvbstsgprttltpassatpglggestatfqtpasttt	504
Db		: : : : : : : :	
Qy	185	QPST-----ESPVLHTSNLPYSYPTINMPWTVSV-----LDTAEFVSDSSSSSS	230
Db		: : : : : : : :	
Qy	505	tpscpcataveesttyhrs--pssrptthfpassattghsekstfhsspdsagttps	562
Db		: : : : : : : :	
Qy	231	SSSSSSGGPLPLPSVQSQHFLFSILPTRASVHLUKST----SDASTPWSSSP-Spl	284
Db		: : : : : : : :	
Qy	563	sabhttsq-----rgesttarispgsteitlpgsttbglseastfyssrspt	613
Db		: : : : : : : :	
Qy	285	----PVSL-----TTTSAPLSVSQTTLPOSSSTPVLPRARETPVTSFORTWTM-	329
Db		: : : : : : : :	
Qy	614	ttlspasmstslvggeesttrsqqgststvspastttpgliseestvysssgpstetv	673
Db		: : : : : : : :	
Qy	330	---SFMTWLHSSQTADLKQSQTPHOEKVIRESKPSLVLPTEST----KAVTNSPLPP	382
Db		: : : : : : : : : : :	
Qy	674	fprsttsvirgeeptthfsrpasthtltftedstts--glteestaafpgspastqtgipa	731
Db		: : : : : : : : : : :	
Qy	383	SLT-----ESSTEQTL-PATSNL-----AQMSPIFTTTLIKTSQPLM	419
Db		: : : : : : : :	
Qy	732	tlctadlgeestffpsgsstcgtklsparsctslvgestpsrlvestettlpgsp--	789
Db		- : : : : : : : :	
Qy	420	TTPQTLSTASLVGPJAVOITAGQLSLTHPEILVPQISTEGGISVERNRVIDVATGL	479
Db		: : : : : : : :	
Qy	790	ttpelsekstfyfsrspdat-----lspattkssgvsee-----sstshs	831
Db		: : : : : : : :	
Qy	480	IPLTSVPTSAKEMTKLGVTAEYSPASRSLSGTSPSPOTTVVYSTAEDLAPKSATFAVO	539
Db		: : : : : : : : : : :	
Qy	832	qpgsthtctafpdsttsglsqepktshssggsteatlpgsttassiigqgsttth-sppg	890
Db		: : : : : : : :	
Qy	540	QSPPTLLSSASVNSCAVNP-----CLHNGECVADNTSRG	573
Db		: : : : : : : :	
Qy	891	dtectlldpddtltsglveastpthsstgslhttltpasstsaq	933
Db		: : : : : : : :	

This Page Blank (uspio)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 14:27:06 ; Search time 18.31 Seconds
(without alignments)
1261.967 Million cell updates/sec

Title: US-09-840-746-1
Perfect score: 4859
Sequence: 1 MSQETVSRVAPMRGEIT.....FPGQYNPSFISDESRDDYF 946

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/6C_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/6D_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4798	98.7	1481	2	US-08-616-844-40
2	4798	98.7	1481	2	US-08-599-654-40
3	4798	98.7	1481	3	US-08-944-868A-40
4	4798	98.7	1481	3	US-08-944-423A-40
5	4798	98.7	1481	3	US-08-944-496-40
6	328	6.8	1257	1	US-08-340-428B-49
7	290.5	6.0	894	3	US-08-362-525-22
8	290.5	6.0	894	3	US-08-971-692-15
9	284.5	5.9	1537	1	US-08-325-267A-2
10	280	5.8	862	1	US-08-325-267A-4
11	279	5.7	2523	1	US-08-185-432-18
12	276.5	5.7	2409	6	5180808-2
13	263.5	5.4	750	4	US-09-165-239A-4
14	254	5.2	2471	1	US-08-185-432-16
15	254	5.2	2471	1	US-08-083-590A-19
16	254	5.2	2471	3	US-08-532-384-19
17	240.5	4.9	2703	1	US-08-185-432-19
18	240	4.9	1193	2	US-08-400-159-10
19	240	4.9	1193	3	US-08-611-729A-10
20	236	4.9	2476	4	US-08-276-967-2
21	233	4.8	1523	4	US-09-182-024A-2
22	231.5	4.8	1010	4	US-08-882-046-7
23	231.5	4.8	1036	4	US-09-068-740A-6
24	231.5	4.8	1187	4	US-09-068-740A-7
25	231.5	4.8	1218	2	US-08-400-159-6
26	231.5	4.8	1218	3	US-08-611-729A-6
27	231.5	4.8	1218	4	US-08-882-046-2

28	231.5	4.8	1218	4	US-09-214-278-7	Sequence 7, Appli
29	231.5	4.8	1218	4	US-09-068-740A-11	Sequence 11, Appl
30	231	4.8	729	3	US-08-872-855-8	Sequence 8, Appl
31	231	4.8	2556	1	US-08-083-590A-20	Sequence 20, Appl
32	231	4.8	2556	3	US-08-532-384-20	Sequence 20, Appl
33	230	4.7	907	3	US-08-783-774-2	Sequence 2, Appl
34	230	4.7	907	5	PCT-US95-04611A-19	Sequence 19, Appl
35	226	4.7	1219	4	US-08-882-046-5	Sequence 5, Appl
36	226	4.7	2556	1	US-08-185-432-17	Sequence 17, Appl
37	225.5	4.6	728	4	US-08-981-392-2	Sequence 2, Appl
38	222.5	4.6	1248	4	US-08-882-046-6	Sequence 6, Appl
39	218.5	4.5	1055	4	US-09-214-278-2	Sequence 2, Appl
40	218.5	4.5	1212	4	US-09-214-278-3	Sequence 3, Appl
41	218.5	4.5	1238	4	US-09-214-278-5	Sequence 5, Appl
42	218.5	4.5	1257	3	US-08-611-729A-8	Sequence 8, Appl
43	217	4.5	1721	3	US-08-928-361B-6	Sequence 6, Appl
44	216.5	4.5	685	3	US-08-872-855-2	Sequence 2, Appl
45	215	4.4	1065	2	US-08-400-159-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-616-844-40
; Sequence 40, Application US/08616844
; Patent No. 5849578
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/616,844
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-616-844-40

Query Match 98.7%; Score 4798; DB 2; Length 1481;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;

Qy 1 MSQETVSRVAPMRGGEITAHWLLTNTSTADVTGSSASYPEGVNASVLTFQSDSTVQS 60
Db 436 MSQETVSRVAPMRGGEITAHWLLTNTSTADVTGSSASYPEGVNASVLTFQSDSTVQS 495
Qy 61 GGSHTALGDRSYSSSSSTSSSLSNAPRGERS----- 94
Db 496 GGSHTALGDRSYSSSSSTSSSLSNAPRGERSLTLEDSREPQALGDSSANAEDRTSGV 555
Qy 95 ----- 94
Db 556 PSLGTHTLATVGTNGERTLRSVTLTNTSMSTTSCEAGSPAAAMPQETEGASLHVNTDDM 615
Qy 95 -----IAGISYQVGRGTAIEQRTSSDHTDHTYLSSTFTKGERALLSITDNS 140
Db 616 GLVSRSLAASSALGVAGISYQVGRGTAIEQRTSSDHTDHTYLSSTFTKGERALLSITDNS 675
Qy 141 SSSDIVESSTSYKISNSHSEYSSSFSHAQTERSNISSYDGEYAQPTESPVLHTSNLPS 200
Db 676 SSSDIVESSTSYKISNSHSEYSSSFSHAQTERSNISSYDGEYAQPTESPVLHTSNLPS 735
Qy 201 YTPPTINPNPNTSVVLDTDAEFVSDSSSSSSSSSSSGPPLPLPSVQSQSHLFFSSILPST 260
Db 736 YTPPTINPNPNTSVVLDTDAEFVSDSSSSSSSSSSSGPPLPLPSVQSQSHLFFSSILPST 795
Qy 261 RASVHLKLTSDASTPMSSSPSPPLVSLTSTSAPLSVSQSTTLPQSSSTPVLPRARETPV 320
Db 796 RASVHLKLTSDASTPMSSSPSPPLVSLTSTSAPLSVSQSTTLPQSSSTPVLPRARETPV 855
Qy 321 TSQTSTMTSFMTMLHSSQADLKQSPTPHQEKVITESKSPSLVSLTSTKAVTTNSPL 380
Db 856 TSQTSTMTSFMTMLHSSQADLKQSPTPHQEKVITESKSPSLVSLTSTKAVTTNSPL 915
Qy 381 PPSLTESSTEQTLPATSTNLQAQSPFTTTILKTSQPLMTTPGTLSTASLVTGPIAVQT 440
Db 916 PPSLTESSTEQTLPATSTNLQAQSPFTTTILKTSQPLMTTPGTLSTASLVTGPIAVQT 975
Qy 441 TAGQLSLTHPEILVPOISTEGGISTERNRNVIVDATTGLIPLTSVPVTSKAKEMTKLGVA 500
Db 976 TAGQLSLTHPEILVPOISTEGGISTERNRNVIVDATTGLIPLTSVPVTSKAKEMTKLGVA 1035
Qy 501 EYSPASRLSGTSPQPTVVYSTAEDLAPKATFAVQSTQSPITLSSASVNSCAVNPCL 560
Db 1036 EYSPASRLSGTSPQPTVVYSTAEDLAPKATFAVQSTQSPITLSSASVNSCAVNPCL 1095
Qy 561 HNGECVADNTRSGYHCRCPSPWQDDGCSVDVNECLSNPCPSTATCNTNQGFTCKCPVG 620
Db 1096 HNGECVADNTRSGYHCRCPSPWQDDGCSVDVNECLSNPCPSTATCNTNQGFTCKCPVG 1155
Qy 621 QLEKGINLVRTFVTEPKRLTEFLNTVVEKHSDLQEVENEITKTLNMCFSALPSYIRSTV 680
Db 1156 QLEKGINLVRTFVTEPKRLTEFLNTVVEKHSDLQEVENEITKTLNMCFSALPSYIRSTV 1215
Qy 681 HASRESNAVVISLQTTFSLASNTVLEFLADRMQKCVNSCKSSAEVCCOLLGSRRIFRAGS 740
Db 1216 HASRESNAVVISLQTTFSLASNTVLEFLADRMQKCVNSCKSSAEVCCOLLGSRRIFRAGS 1275
Qy 741 LCKRKSPECDKTSICTDLGVALCOCKSGYFOFNKMDHSCRACEDGYRLENETCMSCPF 800
Db 1276 LCKRKSPECDKTSICTDLGVALCOCKSGYFOFNKMDHSCRACEDGYRLENETCMSCPF 1335
Qy 801 GLGLNCGNPYQLITVVVIAAAGGGLLLILGIALIVTCCRKNKNDISKLIKFGSGDFQMSPY 860
Db 1336 GLGLNCGNPYQLITVVVIAAAGGGLLLILGIALIVTCCRKNKNDISKLIKFGSGDFQMSPY 1395
Qy 861 AEPKPNPRSQEWGREATEMHENGSKNLQMTDVIYSPTSVRNPELERNGLYPAYTGLPG 920
Db 1396 AEPKPNPRSQEWGREATEMHENGSKNLQMTDVIYSPTSVRNPELERNGLYPAYTGLPG 1455

Qy 921 SRHSCIPPGOYNPSFISDESRRDYF 946
Db 1456 SRHSCIPPGOYNPSFISDESRRDYF 1481

RESULT 2
US-08-599-654-40
; Sequence 40, Application US/08599654
; Patent No. 5882925
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSTICS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/599,654
; FILING DATE: 09-FEB-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-041
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-599-654-40

Query Match 98.7%; Score 4798; DB 2; Length 1481;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;

Qy 1 MSQETVSRVAPMRGGEITAHWLLTNTSTADVTGSSASYPEGVNASVLTFQSDSTVQS 60
Db 436 MSQETVSRVAPMRGGEITAHWLLTNTSTADVTGSSASYPEGVNASVLTFQSDSTVQS 495
Qy 61 GGSHTALGDRSYSSSSSTSSSLSNAPRGERS----- 94
Db 496 GGSHTALGDRSYSSSSSTSSSLSNAPRGERSLTLEDSREPQALGDSSANAEDRTSGV 555
Qy 95 ----- 94
Db 556 PSLGTHTLATVGTNGERTLRSVTLTNTSMSTTSCEAGSPAAAMPQETEGASLHVNTDDM 615
Qy 95 -----IAGISYQVGRGTAIEQRTSSDHTDHTYLSSTFTKGERALLSITDNS 140

Db 916 PPSLTESSTEQTLPATSTNLAQMSPTFTTTILKTSQPLMTTPGTLSTASLVTGPIAVQT 975
QY 441 TAGKQLSLTHPEILVPOISTEGGISTERNRVIIVDATGTLPLTSVPTSAKEMTKLGVT 500
Db 976 TAGKQLSLTHPEILVPOISTEGGISTERNRVIIVDATGTLPLTSVPTSAKEMTKLGVT 1035
QY 501 EYSPASRLSGTSPQTTVVVSTADLAPKSAATFAVQSSSTQPTTLSSASVNSCAVNPCL 560
Db 1036 EYSPASRLSGTSPQTTVVVSTADLAPKSAATFAVQSSSTQPTTLSSASVNSCAVNPCL 1095
QY 561 HNGECVADNTRSGYHCRCPSPWQDDCSVDVNECLSNPCPSTATCNTQSGFICKCPVG 620
Db 1096 HNGECVADNTRSGYHCRCPSPWQDDCSVDVNECLSNPCPSTATCNTQSGFICKCPVG 1155
QY 621 QLEKGCINLVRTFTVTEFKLARTFLNTTVEKHSDLQVEVEITKTLNMCFSALPSYIRSTV 680
Db 1156 QLEKGCINLVRTFTVTEFKLARTFLNTTVEKHSDLQVEVEITKTLNMCFSALPSYIRSTV 1215
QY 681 HASRESNAVVISQTTFTSLASNTVTLFDLADRMOKCVNSCKSSAEVQQLGSRRIFRAGS 740
Db 1216 HASRESNAVVISQTTFTSLASNTVTLFDLADRMOKCVNSCKSSAEVQQLGSRRIFRAGS 1275
QY 741 LCKRKSPCKDTSICTDLGVALCCKSGYFQFNKMDHSCRACEDGYRLENETCSCPF 800
Db 1276 LCKRKSPCKDTSICTDLGVALCCKSGYFQFNKMDHSCRACEDGYRLENETCSCPF 1335
QY 801 GLGLNCGNPQYQITVVVIAAGGGLLILGIALIVTCCRKNKNDISKLIFFKSGDFQMSPY 860
Db 1336 GLGLNCGNPQYQITVVVIAAGGGLLILGIALIVTCCRKNKNDISKLIFFKSGDFQMSPY 1395
QY 861 AEPKPNRQSEGREATEMHENGSTKNLQMTDVVYSPTSVRNPELERNGLYPAYTGLPG 920
Db 1396 AEPKPNRQSEGREATEMHENGSTKNLQMTDVVYSPTSVRNPELERNGLYPAYTGLPG 1455
QY 921 SRHSCLFPGQYNPSFSDSRRRDYF 946
Db 1456 SRHSCLFPGQYNPSFSDSRRRDYF 1481

RESULT 4
US-08-944-423A-40
; Sequence 40, Application US/08944423A
; Patent No. 6020463
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS 95
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,423A
; FILING DATE: 06-OCT-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: JUN-07-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995

; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-944-423A-40

Query Match 98.7%; Score 4798; DB 3; Length 1481;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;

QY 1 MSQTEVSRVAPMRGGEITAHLLTNTSTTSADVGTGSSASYPEGVNASVLTQFSDSTVQS 60
Db 436 MSQTEVSRVAPMRGGEITAHLLTNTSTTSADVGTGSSASYPEGVNASVLTQFSDSTVQS 495
QY 61 GGSHTALGDRSYSESSSTSSSESLNSAPRGERS- ----- 94
Db 496 GGSHTALGDRSYSESSSTSSSESLNSAPRGERS- ----- 555
QY 95 ----- 94
Db 556 PSLGTHLTATVGTNGERTLSRVTLTNTSMSTTSGEAGSPAAAMPQETEGASLHVNVTDDM 615
QY 95 -----IAGISYGVQVRGTAIEQRTSSDHDHTYTLSTFTTKGERALLSITDNS 140
Db 616 GLVSRSLAASALGVAGISYGVQVRGTAIEQRTSSDHDHTYTLSTFTTKGERALLSITDNS 675
QY 141 SSSDIVESSTYIKISNSHSEYSSFSHAQTERNSYSDGEYAQPTESPVLHTSNLPS 200
Db 676 SSSDIVESSTYIKISNSHSEYSSFSHAQTERNSYSDGEYAQPTESPVLHTSNLPS 735
QY 201 YPTINMPNTSVVLDTDAEFVSDSSSSSSSSSSSGGPPPLPSPVQSQSHLFSILPST 260
Db 736 YPTINMPNTSVVLDTDAEFVSDSSSSSSSSSSSGGPPPLPSPVQSQSHLFSILPST 795
QY 261 RASVHLLKSTSDASTPMWSSSPPLPVSLTSTAPLSVSQTTLPQSSSTPVLPRARETPV 320
Db 796 RASVHLLKSTSDASTPMWSSSPPLPVSLTSTAPLSVSQTTLPQSSSTPVLPRARETPV 855
QY 321 TSFQSTMTSFTMTLHSSQTDADLKQSQTPHQEKVITESKSPSLVSLPTESTKAVTTNSPL 380
Db 856 TSFQSTMTSFTMTLHSSQTDADLKQSQTPHQEKVITESKSPSLVSLPTESTKAVTTNSPL 915
QY 381 PPSLTESSTEQTLPATSTNLAQMSPTFTTTILKTSQPLMTTPGTLSTASLVTGPIAVQT 440
Db 916 PPSLTESSTEQTLPATSTNLAQMSPTFTTTILKTSQPLMTTPGTLSTASLVTGPIAVQT 975
QY 441 TAGKQLSLTHPEILVPOISTEGGISTERNRVIIVDATGTLPLTSVPTSAKEMTKLGVT 500
Db 976 TAGKQLSLTHPEILVPOISTEGGISTERNRVIIVDATGTLPLTSVPTSAKEMTKLGVT 1035
QY 501 EYSPASRLSGTSPQTTVVVSTADLAPKSAATFAVQSSSTQPTTLSSASVNSCAVNPCL 560
Db 1036 EYSPASRLSGTSPQTTVVVSTADLAPKSAATFAVQSSSTQPTTLSSASVNSCAVNPCL 1095
QY 561 HNGECVADNTRSGYHCRCPSPWQDDCSVDVNECLSNPCPSTATCNTQSGFICKCPVG 620
Db 1096 HNGECVADNTRSGYHCRCPSPWQDDCSVDVNECLSNPCPSTATCNTQSGFICKCPVG 1155
QY 621 QLEKGCINLVRTFTVTEFKLARTFLNTTVEKHSDLQVEVEITKTLNMCFSALPSYIRSTV 680
Db 1156 QLEKGCINLVRTFTVTEFKLARTFLNTTVEKHSDLQVEVEITKTLNMCFSALPSYIRSTV 1215

Qy 681 HASRESNAVVISLQTTFFSLASNVTLFDLADRMQKVCNSCKSAEVCQLLGSORRIFRAGS 740
|||||
Db 1216 HASRESNAVVISLQTTFFSLASNVTLFDLADRMQKVCNSCKSAEVCQLLGSORRIFRAGS 1275
|||||
Qy 741 LCKRKSPECDKDTISCTDLGVALCOCKSGYFQFNKMDHSCRACEDGYRLENETCMSCPF 800
|||||
Db 1276 LCKRKSPECDKDTISCTDLGVALCOCKSGYFQFNKMDHSCRACEDGYRLENETCMSCPF 1335
|||||
Qy 801 GLGGLNCNPNYQLITVIVIAAAGGGLLLILGIALIVTCCRNKNDISKLIKFSQDFQMSPY 860
|||||
Db 1336 GLGGLNCNPNYQLITVIVIAAAGGGLLLILGIALIVTCCRNKNDISKLIKFSQDFQMSPY 1395
|||||
Qy 861 AEYPKNPRSQEWGREAIEMHENGSTKNLQMTDVYYSPTSVRNPELERNGLYPAYTGLPG 920
|||||
Db 1396 AEYPKNPRSQEWGREAIEMHENGSTKNLQMTDVYYSPTSVRNPELERNGLYPAYTGLPG 1455
|||||
Qy 921 SRHSCIFPGQYNPSPISDESRRRDYF 946
|||||
Db 1456 SRHSCIFPGQYNPSPISDESRRRDYF 1481
|||||

RESULT 5
us-08-944-496-40
; Sequence 40, Application us/08944496
; Patent No. 6124433
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF INVENTIONS: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,496
; FILING DATE: 06-OCT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
us-08-944-496-40

Query Match 98.7%; Score 4798; DB 3; Length 1481;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;
Qy 1 MSQETVSRSVAPMRGGEITAHWLLTNTTSSADVTGSSASYPEGVNASVLTQFSDSTVQS 60
|||||
Db 436 MSQETVSRSVAPMRGGEITAHWLLTNTTSSADVTGSSASYPEGVNASVLTQFSDSTVQS 495
|||||
Qy 61 GGSHTALGDRSYSESSSTSSSESLSNSAPRGERS----- 94
|||||
Db 496 GGSHTALGDRSYSESSSTSSSESLSNSAPRGERSLTEDSREPGQALGDSSANAEDRTSGV 555
|||||
Qy 95 ----- 94
|||||
Db 556 PSLQHTTLATVTGNGERTLRSVTLTNTSMSTTSGBAGSPAAAMPQETEGASLHVNVTDMM 615
|||||
Qy 95 -----IAGISYGQVGTATJEQRTSSDHTDHTYLSSTFTTGERALLSITONS 140
|||||
Db 616 GLVSRSLAASALGVAGISYGQVGTATJEQRTSSDHTDHTYLSSTFTTGERALLSITONS 675
|||||
Qy 141 SSSDIVESSTYIKISNSHSEYSSFSHAQTERSNISSYDGEYAQPSTESPVLTHTSNLPS 200
|||||
Db 676 SSSDIVESSTYIKISNSHSEYSSFSHAQTERSNISSYDGEYAQPSTESPVLTHTSNLPS 735
|||||
Qy 201 YTPRTINMPNTSVVLDTDAEFVSDSSSSSSSSSSSSSGPPPLPSPVSQSHHLFSSILPST 260
|||||
Db 736 YTPRTINMPNTSVVLDTDAEFVSDSSSSSSSSSSSSSGPPPLPSPVSQSHHLFSSILPST 795
|||||
Qy 261 RASVHLKSTSDASTPWSSSPSPVSLTSTTSAPLSVSQTTLPQSSSTPVLPRARETPV 320
|||||
Db 796 RASVHLKSTSDASTPWSSSPSPVSLTSTTSAPLSVSQTTLPQSSSTPVLPRARETPV 855
|||||
Qy 321 TSFQSTMTSEMTMLHSSQTDADLKQSQSTPHOEKVITESKPSLSVLPSTESKAVTNSPL 380
|||||
Db 856 TSFQSTMTSEMTMLHSSQTDADLKQSQSTPHOEKVITESKPSLSVLPSTESKAVTNSPL 915
|||||
Qy 381 PPSLTESSTEOTLPATSTNLAQMSPTFTTILKTQSPLMTTPGTLSSTASLTGPIAVQT 440
|||||
Db 916 PPSLTESSTEOTLPATSTNLAQMSPTFTTILKTQSPLMTTPGTLSSTASLTGPIAVQT 975
|||||
Qy 441 TAGKQLSLTHPEILVPOISTEGGISTERNRNVIVDATTGLIPLTSVPTSASAKEMTTKLGVTA 500
|||||
Db 976 TAGKQLSLTHPEILVPOISTEGGISTERNRNVIVDATTGLIPLTSVPTSASAKEMTTKLGVTA 1035
|||||
Qy 501 EYSPASRLGTSPPQTTVWSTAEADLPKSAFVQSTQSTPTTLSSASVNSCAVNFCL 560
|||||
Db 1036 EYSPASRLGTSPPQTTVWSTAEADLPKSAFVQSTQSTPTTLSSASVNSCAVNFCL 1095
|||||
Qy 561 HNGECVADNTSRGYHCRCPSPWQGGDCSDVNECLSNPCPSTATCNCNTOGSGFICKCPGY 620
|||||
Db 1096 HNGECVADNTSRGYHCRCPSPWQGGDCSDVNECLSNPCPSTATCNCNTOGSGFICKCPGY 1155
|||||
Qy 621 QLEKGICNLVRTFVTEFKLKRFTLNTTVEKHSDLQOEVEINEITKTLNMCFSALPSYIRSTV 680
|||||
Db 1156 QLEKGICNLVRTFVTEFKLKRFTLNTTVEKHSDLQOEVEINEITKTLNMCFSALPSYIRSTV 1215
|||||
Qy 681 HASRESNAVVISLQTTFFSLASNVTLFDLADRMQKVCNSCKSAEVCQLLGSORRIFRAGS 740
|||||
Db 1216 HASRESNAVVISLQTTFFSLASNVTLFDLADRMQKVCNSCKSAEVCQLLGSORRIFRAGS 1275
|||||
Qy 741 LCKRKSPECDKDTISCTDLGVALCOCKSGYFQFNKMDHSCRACEDGYRLENETCMSCPF 800
|||||
Db 1276 LCKRKSPECDKDTISCTDLGVALCOCKSGYFQFNKMDHSCRACEDGYRLENETCMSCPF 1335
|||||
Qy 801 GLGGLNCNPNYQLITVIVIAAAGGGLLLILGIALIVTCCRNKNDISKLIKFSQDFQMSPY 860
|||||
Db 1336 GLGGLNCNPNYQLITVIVIAAAGGGLLLILGIALIVTCCRNKNDISKLIKFSQDFQMSPY 1395
|||||
Qy 861 AEYPKNPRSQEWGREAIEMHENGSTKNLQMTDVYYSPTSVRNPELERNGLYPAYTGLPG 920
|||||
Db 1396 AEYPKNPRSQEWGREAIEMHENGSTKNLQMTDVYYSPTSVRNPELERNGLYPAYTGLPG 1455
|||||

Qy 921 SRHSCIFPGQNPFSFIDSESRDDYF 946
Db 1456 SRHSCIFPGQNPFSFIDSESRDDYF 1481

RESULT 6

US-08-340-428B-49
; Sequence 49, Application US/08340428B
; Patent No. 5648465
; GENERAL INFORMATION:
; APPLICANT: MARGOLIS, Richard U.
; APPLICANT: RAUCH, Uwe
; APPLICANT: MARGOLIS, Renee K.
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
; TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,428B
; FILING DATE: 14 No. 5648465ember 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/922,911
; FILING DATE: 03 August 1992
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: Margolis-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-340-428B-49

Query Match 6.88; Score 328; DB 1; Length 1257;
Best Local Similarity 22.0%; Pred. No. 2.9e-14;
Matches 155; Conservative 107; Mismatches 255; Indels 186; Gaps 31;
Qy 1 MSOTETYSRSVAPMRGGEITAHWLLTNSTTSADVTGSSA-----SYPEGVNASVLTOF 53
Db 418 LTRFQASQETLASTPGPTLASWLLTGVTSTGVPSFSLGVDMEETTPSG-----TQV 471
Qy 54 SDSTVQSGSHALTGDRSYSESSSTSSSESLNSAPRGERSIAGISYGVGRGTAIEQRTS 113
Db 472 APTPTMRGRFGLNGRHFQOQG--PEDQLLEAAEASQPTTLEVTADHMGPSAATALE 529
Qy 114 SDHTDHTYLSFTFKGERALLSITDSSSDIVESSTSYKISNSSHSEYSSFSHAQTER 173
Db 530 SDQSHSPW-----ATL-----TNEVDVPGAGS-----LGRSLPESRKNSPSLIS 570
Qy 174 SNISSYD-----GEVAQSTESPVLHTSNLPSYPTINPNPNTSVVLDTADEFVSDSS 226
Db 571 STVPSTSTPGLKPGADEATGVKSAIHPPLWPS-EPAV-----PSSIPSEALS 618

Qy 227 SSSSSSSSSSGPPLPLPSVSQSHHLF-----SSILPSTRASVHLLKSTSDASTPWSSSPS 282
Db 619 AVSLQASPGDGDSPDFPIVAMLRAPKLLPHSLTPVNV-----S 657
Qy 283 PLPVSLTTSAPLSVSQTTLPQSSSTPVLPRARETPTVTSQTSMTWTFMTLHSSQTAD 342
Db 658 PIPLS-----PASPLPSSYPPEQAVRPVSFGAEDPETPFQT-TMAAPGEASHGSPDAD 709
Qy 343 -----LKQSTPHQEKVITESKSPSLVSLPTES--TKAVTNSPLPPLPSTESSTEQ-- 391
Db 710 SIEIEGISSMOATKHP-----ISCPWASLSSNVTVN-PVPDAGILGTESGV 756
Qy 392 -TLPATSTNLAQMSPTFTTILKTSQPLMTTPGTLSTSLTASLVTGPIAVQTTAGKQLSLTH 450
Db 757 LDLPGSPTSDQAT---VDMVLATWLPL---PGHGLDTSQST-PMEAH---GYTMSVE- 805
Qy 451 PEILVQISTEGGISTERNRVIVDATTCGLIPLTSVPTSAKE-----MTTKLGVTFA---EY 502
Db 806 -----PTVALEGGATKDP---MEATMDVVPSTVDATSGSEPKSSISSTHVVVTAAGDQG 856
Qy 503 SP-----ASRSLGTSPS---PQTTVVSTAEDLAPKATFAVQ---SSTQSP 542
Db 857 TPTLTPTSSEGVVAQESLGLTSLPSHPWSSSLASSMDEVASVSGEPTRLWDIPSTLIP 916
Qy 543 TTLSSAS-----VNSCAVNPCLHNGECVADNTSRGYHCR 577
Db 917 VSLGLDESCLKVAESPCLEGFEWEEVAGSQEDPTDPCENNPFCLHGGTCRTNGTM--YGCS 974
Qy 578 CPPSWQGDGDSVYNECLSNPCPSTATYCNNTQGSFICKPCVGY 620
Db 975 CDQGYAGENCEIDDDCLCSPCENGCGTCIDEVNGFICLCPLSY 1017
RESULT 7
US-08-362-525-22
; Sequence 22, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KLIS, FRANCISCUS M.
; APPLICANT: SCHREUDER, MAARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92202080.5
; FILING DATE: 08-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92203899.7
; FILING DATE: 14-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/01763
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.

```

; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 213289/77020(V)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 894 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-362-525-22

Query Match          6.0%; Score 290.5; DB 3; Length 894;
Best Local Similarity 23.8%; Pred. No. 6.9e-12;
Matches 154; Conservative 96; Mismatches 244; Indels 153; Gaps 24;

QY 38 SASYPEGVNAS-----VLQPSDSSTVSGGSHGTALGDRSYSESSSTSSSELN----SSAP 89
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 234 SVTLPGDTTVSDDFEGYVYSFDDLLSQS--NCTVPDPNSYAVSTTTTTTPEWTGFTST 291
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 90 RGRSIAGISGVGRGTAIEORTSSDH-----TDHTYLSSTFTKGERALLSIT-DNSSSD 144
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 292 TEMTVTGTNGVPTDETVIVIRTFSEGLISTTEPMTGTFTSTEVTTITGTNGQPTD 351
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 145 ----IVSSSTSVKISNSSHSEYSFSFHAQPERGNSISYDGE-----182
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 352 ETVIVIRPTSEGLISYITEPWTGFTSTSTEMVTGTNGQPTDETVIVIRTFPSEGLV 411
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 183 -----YAQPSTE--SPVLHSTNLPSPYPTINMPNTSVVLDTDAEFVSDSSSSSSSS 231
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 412 TTTPEMTGTFTSTSTEMSTVTGTNGLPT-----DETIVVWKTPTTAISSLSLSSSSG 464
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 232 --SSSSSGPPLP-----SVSQSHLFSSILPSTRASVHLKS-----TSDAS 274
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 465 QITSITSSRRPIITFPYPSNCTSVISSSVSSSVTSLFTSPVISSSVSSSTTTSTSI 524
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 275 TPWSSSSPLPVSLTTS-----TSAPLSVSQTTL-----POSSSTPVLPR----- 315
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 525 FSESSKSVIPTSSSTSGSSESETSSAGSVSSSFSESSKSPYSSSLPLVTSATTS 584
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 316 RET-----PVTSFQTSNMTSPWMLHSSQTADLKSQTPHPEKVITESKSPSLVSLPTES 370
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 585 QETASSLPPATTTKTSEQTTLTV-----TSCSHVCTESISPAIVSTATVT 631
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 371 TKAVTT--NSPLPPLSTES-----STEQTLPAT--SNLAQMSPTFTTILKTQSP--LM 419
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 632 VSGVTYTEYTWCPISITTEYTKQKGTTEQFTTKQTVVTVISSCEDVCSKTASPAIVS 691
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 420 TTPGTLSTASLVPGLPAVQITACKQ---LSLTHPELV-----455
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 692 TSTATNGVTEYTWCPISISTEYRQQTLLVTVTSCSGVCSFETSPASPAIVSTATATVNDV 751
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 456 -----PQISTEGGISTERNRIVDATTC--LIPLTSVPTSAKEMTTLKGVTAEYSPAS 506
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 752 VTVPYTRPQTANEEYSVSKNSATGETTTNTLAAETTTNVAETITNTCAAETKTVT 811
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 507 RSLGTSPTQPTVYVTAEDLAPKSAFPAVQSSQTSPPTLSSSASVNS 553
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 812 SSLSRNHAETQTATSDTVIGHSSSVSVSVSETGNTKSLTSSGLSTMS 858
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
US-08-971-692-15
; Sequence 15, Application US/08971692
; Patent No. 6114147
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Immobilized proteins with specific binding
; TITLE OF INVENTION: capacities and their use in processes and products.
; NUMBER OF SEQUENCES: 40

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,692
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 894 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-971-692-15

Query Match          6.0%; Score 290.5; DB 3; Length 894;
Best Local Similarity 23.8%; Pred. No. 6.9e-12;
Matches 154; Conservative 96; Mismatches 244; Indels 153; Gaps

QY 38 SASYPEGVNAS----VLTQFSDSTVQSGGSHALGDRSYSESSSSSSSESLSN---SSAP 89
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 234 SVTLQDGTTVDDDFEGVYGFDDQLSGS---NCIVDPSPNVAVSTTTTTTTPWTGTFST 291

QY 90 RGRSIAIGYGVGRGTAIBQRTSSDH----TDHTVLSSTFTKGERALLASIT--DNSSSD 144
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 292 TEMTVTGTNGVPTDEIVIRIPTSEGLISTTTEPWGCTGTFSTSEVTITGTNGQPTD 351

QY 145 -----IVESSTSYKISNNSHSEYSSFSHAQTERNSISSYDGE-----182
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 352 ETIVIRIPTSEGLISTTTEPWGCTGTFSTSEMTVTGTNGQPTDEIVIRIPTSEGLV 411

QY 183 -----YAPQSTE--SPVLHTSNLPSVTPINPNMTSVWLDTDAEFVSDSSSSSSSS 524
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 412 TTTTPWTGTFSTSEMTVTGTNGLPT-----DETIVVVKTPPTAISLSSLSSSSG 524

QY 232 --SSSSSGGPPPLP-----SVSOSHHLFSSILPSTRASVHLKS-----TSDAS 274
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 465 QITSSITSSRPITPPYPNGTSVISSSVISSSVTSSLFTSSPVISSSVISSSTTTSTSI 524

QY 275 TPWSSSPSPPLVSLTTS-----TSAPLSVSQVTL-----POSSSTPVLPR---315
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 525 FSESSKSVIPTSSSTSGSSESETSAGSVSSSSSFISSESSKSPYSSSGLPLVTSATTS 584

QY 316 RET-----PVTSQTSMTSFMWLHSSQADLKSQSTPHQKVIYTESKPSLVSLPTES 370
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 585 QETASSLPAPATTKTSQVTLTV-----TSCSHVCFTESISPAIVSTATVTT 631

QY 371 TKAVTT--NSPLPPLSTES-----STEQLPAT--STNLAOMSPTTTTTLTKTSOP--LM 419
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 632 VSGVTTEYTWCPDISTETTKQKTEQQTETTKQTVVTVTSSCEDSDVCSKTASPAIVS 691

QY 420 TTRGTSTASLVTGPIAVQTAGKQ----LSLTHPEILV-----455
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 692 TSPATINGVTTEYTWCPDISTESRQQTLLVTVTSCSGVCSSETASPAIVSTATATVNDV 751

QY 456 -----POISTEGGLISTERNRVIVDATTG--LIPLTSPVPTSAKEMTKLGVTAEYSPAS 806
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 752 VTVPYTPRQPTANEEYSYKKNKSNATGETTTNTLAAETTTNTVAEATITNTGAAETKTVT 811

QY 507 RSLGTSFSPQPTVVSTAEDLAPKSAFVAQSSSTQSPPTLLSSASVNS 553
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 812 SSLRSNHAEQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTWS 858

RESULT 9
US-08-325-267A-2
; Sequence 2, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI

```


LENGTH: 862 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-267A-4

Query Match 5.8%; Score 280; DB 1; Length 862;

Best Local Similarity 23.7%; Pred. No. 3.5e-11; Mismatches 232; Indels 144; Gaps 24;

38 SASYPEGVNAS-----VLTQSDSTVSGSGSHATGDRSYSESSSSSESLN---SSAP 89
Db 234 SVTLPGDTGTVDDPEGVYVDFDLSQS--NCTVPDPSNVAVSTTTTTPWTGTFTSTS 291
QY 90 RGRSIAGISGVGRGTALQRTSDHTDHTYLS-----TFTKGRALLSIT-DNSSS 142
Db 292 TEMTIVTGTNGVPTDEVIIVIRPT--TASTIIITTPWTGTFTSTSTEMTIVTGTNGQP 349
QY 143 SD----IVESSTSVIKISNSSHSEYSSFSHAQTERSNISSVDGEYAQPSTESPVLHTSNL 198
Db 350 TDETVIVIRPTSGLVTTTTEPTGTFTSTSTEMSVTG-----TNGL 393
QY 199 PSYPTINMPNTSVYLDTADEFVSDSSSSSSSS--SSSSSGGPPPLP-----SVSQS 249
Db 394 PT-----DETIVVKTPTTAISSLSSSSSSGOITSSITSSRPIITPFYPSNGTSVISS 446
QY 250 HHLFSSILPSTRASVHLKS-----TSDASTPWSSSPSPPLVSLTTS-----TSAP 295
Db 447 SVISSSVTSSILFTSPVSSVSSSTTTSTSPFSSKSSVIPTSSSTSGSSESETSSA 506
QY 296 LSVSOTTL-----POSSGFPVLPRA---RET----PVTSFOTSTMTSPMTLHS 337
Db 507 GSVSSSFISESSKSTYSSSLPLVTSATSOETASSLPPTATTKTSEQTILTV--- 563
QY 338 SQTADLKSQTPHOEKVITESKPSLSLPTSTKAVTT--NSLPPSLTES-----ST 389
Db 564 -----TSCESHVCTESISPAIVSTATVTVSGVTTTEVTWCPISTTTTKTKGTT 613
QY 390 EQTLPAT-SNLAQMSPTFTTILKTQSP--LMTPTGLSTASLVGPIAVQTAKQ- 445
Db 614 EQTETTKQTQTVVTISSCEDVCSTKSPAIVSTSTATINGVTTEYTWCPISPTESRQ 673
QY 446 ---LSLTHPELV-----POISTEGGSTERNRVID 474
Db 674 TTVTVTSCSGVCSEFASPAIVSTATATVNDVVTVPTWRPQTANESVSSKMSATGE 733
QY 475 ATTG-LIPLTSVPTSASAKEMTKLGVTAEYSPASRLGTSPTQTVTVSTAEDLAPKSAIF 533
Db 734 TTTNTLAAETTTNTVAAETIINTGAAETKTVTSSLSRNSHAETQTASATDVICHSSVV 793
QY 534 AVQSTOSPTTLSSASVNS 553
Db 794 SVSETGNTKSLTSSGLSTMS 813

RESULT 11

US-08-185-432-18

Sequence 18, Application US/08185432

Patent No. 5750652

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon

APPLICANT: Busseau, Isabelle

APPLICANT: Diederich, Robert J.

APPLICANT: Xu, Tian

APPLICANT: Matsuno, Kenji

TITLE OF INVENTION: DELTA PROTEINS, NUCLEIC ACIDS, AND

TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2523 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-18

Query Match 5.7%; Score 279; DB 1; Length 2523;

Best Local Similarity 23.5%; Pred. No. 2e-10;

Mismatches 75; Conservative 35; Mismatches 94; Indels 118; Gaps 12;

QY 550 SVNSCAVNPCLHNGECVAD-----NTSR----- 572
Db 790 NINECSNPNCLHGTCTDDVAGYKCNMCLPYTGAICAVLAPCAGSPCKNGRCKESDF 849
QY 573 -GYHCRPPSQGDDCSVDVNECLSNPCPSTATCNCNTQGSFICKCPVGYOLEKICNLVR 631
Db 850 ETFSCGPPGQGTCEIDMNECVNRPCRGATCQNTGYSKCKCKPGY----- 898
QY 632 TFVTEFKRLTFLVTVEKHSDLQVE-----NETTKLNMCFALPSVIRSTVHASR 684
Db 899 -----TGRNCEMDIDCQPNPCNCHNGSCSDGINMFCNCPAGFRGP----- 939
QY 685 ESNNAVISLQTTFSLASNVTLFDLADRMQKV-NSCKSSAEVQCLLAGSQRIFR---AGS 740
Db 940 -----KCEEDINECASNPCKNGANCTDCVNSYTCCTCQPGFSGI 977
QY 741 LCKRKSPECCKDTSIC-----TDLGV--ALCOCKSGYGFQFNKMDHSCRACEDGYRLENET 794
Db 978 HCESNTPDCTE--SSCFNGGTCIDGINTFTTCQCPPG-FTGSYCOHDINECDKPKCLNGGT 1034
QY 795 CM-----SCPFGLGLGNCN 809
Db 1035 QDSYGYTKCTCPQGYTGLNCN 1057

RESULT 12

5180808-2

Patent No. 5180808

APPLICANT: RUOSLAHTI, ERKKI I.

TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID

SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN

ANTIBODIES, AND METHODS OF DETECTING THE SAME

NUMBER OF SEQUENCES: 4

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/441,179

FILING DATE: 27-NOV-1989

SEQ ID NO:2

LENGTH: 2409

5180808-2

```
Query Match          5.7%; Score 276.5; DB 6; Length 2409;
Best Local Similarity 20.0%; Pred. No. 2.8e-10;
Matches 182; Conservative 104; Mismatches 280; Indels 343; Gaps 36;

Qy 2 SQTETSRVAPM-----RGEITAHLLTNSTTSADVTGSSASYPEGVNASVLTQF 53
Db 1315 NRMENVAKEGVLVSQTDIFEGSGSVT-----STTLIELSDTGA--EGPTVAPLPFS 1365
Qy 54 SD-----STVQSGSHTALGDRSYSESSSTSSSESLSNSA----- 88
Db 1366 TDIGHPONQVRMAEETQTRPOTITEQDNKNKSSTAETNETTSSDFFLARAYFFEMAK 1425
Qy 89 -----PRGERSIAGISGVQVRGTAIEQRTSSDHTD--HTYLSSTFTKGER 131
Db 1426 EFVTSAPKPSDLYVEPSEGS-----GEV-----DIVDSFHTSATQATQRES 1468
Qy 132 ALLSITONS-----SSSDIVESSTSYTKINSNHSSEYSSFSHAQTER-SNissy--- 179
Db 1469 STTFVSDGSLKHPVPKSAKAVTAGPPTVSVMLPLHSEQNKSPDPTSLSNVTYSERS 1528
Qy 180 DGEY-----AOPS-----TESPVLHTSNLPSVT-- 202
Db 1529 TDGSGQDRPREFEDSTLKNRKKPTENIIDLDKEDKDLITITESTILEI--LPELTSD 1586
Qy 203 --PTINNPTSV-----LQDAEFVSDSSSSSSSSSSSS----- 236
Db 1587 KNTIIDIDHTKPVYEDILGMQTDIDTEVPSEPHDSNDSNDSTQVQEIYEAANLSLITE 1646
Qy 237 ---SGPPLPLPSVSQSHHLLFSSILPSTRASVHLLKSTSDASTPSSPSPPLVSLTSTS 293
Db 1647 ETEFGSADVLASTYQATH--DESMYEDRQLDHMGHFHTTGIPAPSTETELDVLLPTATS 1705
Qy 294 APL--SVSQTLTPOSSSTPVLPRARETEVTSFOTSTMTSFMTMLHSSOTADLKQSTPHQE 352
Db 1706 LPIPKSATVIPETEGIKAKAKALD---DMFESTLSDGQAIADQSEIILITLQGFERTQE 1762
Qy 353 KVITESKPSLSVLPTL-----STKAVTNTSPPLPSLFTESTEQTLPATSNLQMSPTFT 408
Db 1763 EY--EDKKHAGPSQPEFSSGAEALVDHTPY-----LSIATTHLMDQSVTEV 1808
Qy 409 TTILKTSQPLMTPCTL--SSTASL-----VTGPVAVQTTAGKQLSLTHPELLVPOISTEG 462
Db 1809 PDVMEGNSPPYYTDTTAVSTFAKLSSQTPSSPLTIY--SGSEAS--GHTET--POPSALP 1863
Qy 463 GI-----STERNRIVDAT-----TGLIPLTSVPVTSKAKEMTKLG 497
Db 1864 GIDVGSSVMSPQDSFKIEIHNIEATFKPSESEYLIHITEPPLSPDFTKLEPSEDDGKPELL 1923
Qy 498 VTAEYSP-----ASRLSGTSPSPQT--TVVSTAEGL----- 526
Db 1924 EEMEASPTELIAVEGTEILODFQNKDQVSGEAIKMFPTIKTPEACTVITTADEIELEG 1983
Qy 527 -----APKSATFAVOS-----STQSPFTLSSSASVN----- 552
Db 1984 ATOMPHSTASATYGVGAEVVPVWMLSPQTSERPTLSSSPINPETQAAALIRQDQSTIAASE 2043
Qy 553 ----- 552
Db 2044 QQVAAARILSDQNTVNPVENTEVATPPPSLLETSHNETDFLIGINESVEGTALYLPGP 2103
Qy 553 -SCAVNPCLHNGECVADNTSRGYHCRPPSQWQGDSCVDVNECLSNPCPSTACNNPQGS 611
Db 2104 DRCKMNPCLNGGTCYPTETS--YVCTCVPGYSGDQCELDLDFDECHSNPCRNCGATCVDFGNT 2161
Qy 612 FICKCPVGY 620
Db 2162 FRCLCLPSY 2170
```

RESULT 13
US-09-165-239A-4
; Sequence 4, Application US/09165239A

```
; Patent No. 6344554
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, ALEXANDER
; APPLICANT: BRAUN, BURKHARD R
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES FROM CANDIDA
; TITLE OF INVENTION: ALBICANS ENCODING POLYPEPTIDES ASSOCIATED WITH FILAMENTOUS
; FILE REFERENCE: 220022000700
; CURRENT APPLICATION NUMBER: US/09/165,239A
; PRIOR FILING DATE: 1998-10-01
; PRIOR FILING DATE: 1997-12-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 750
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-165-239A-4
```

```
Query Match          5.4%; Score 263.5; DB 4; Length 750;
Best Local Similarity 22.4%; Pred. No. 4e-10;
Matches 153; Conservative 100; Mismatches 238; Indels 193; Gaps 29;

Qy 44 GYNASVLTQFSDSTVQSGGSHALGDRSYSESSSTSSSESLSNSAPRGER----SIAGIS 99
Db 80 GYNAGFAAFVVSNAKK-----LSDGSGIDCNFKSDSSVOLNLAFGKKVKQLSITGTG 133
Qy 100 XGQVR-----GTAIEQRTSSDHTHTYLSSTFTKGERALLS-----ITD---NSSSSDIVE 147
Db 134 YSDISLLGNVANPPEWSAS-----LKVKAIVKGGKCLPSPGFRIVTDFESNCPEFDAIK 187
Qy 148 ----SSTSYTKISNSSHS----EYSSFSHAQT-----ERSNIS----- 177
Db 188 QFGSSQIIVKNAVSNAGTFDASALENAQVKAFFAKPAKRELDSEELNSDGVTHSKRTLK 247
Qy 178 -----SYDGEYAAQPTESPVLYHTSNLPSYTTINMPTNSVVLDTDA 218
Db 248 LLLGLLKKVTGGCDTLQFCWDCQCDTPSPSTTTVTSTSSAPSTSPSSAPSTTTVTSSS 307
Qy 219 EFVSDSSSSSSSSSSSSSGP-PLPLPSVSQSHHLLFSLIPSTRASVHLLKSTSDASTPW 277
Db 308 PVTSPSSPVPETTTVTSSVPETTPESSAPETTTVTSSVPETTPESSAPETTPESSAPE 367
Qy 278 SSSPSPVLSLTTSTASPLSVSQTLPQSSS---TPVLPRARETPTVTSFOTSTM----- 328
Db 368 SSVPESSAPETTPESSAP-----ESSVPSSAPETETETPTTAHLTTTITA-OTTIVITVTS 422
Qy 329 -----TSFMTMLHSSQTAADLKS-----QSTPHQEKVI 355
Db 423 CSNNACSKTEVTGCVVVTSEDTIYTTFCPLTETTPVPSSVDSTSVTSAPETTP--ESTA 480
Qy 356 TESKPSPLSVLPTSTKAVTNSPLPPLSLT---ESSTEQTLPATSNLQMSPTFTTIL 412
Db 481 PESSAPE--SSAPESS--APVTETPGVSTVTEQSKTIVTITSCNNACSESKVTTGVVV 537
Qy 413 KTSO-----PLM--TPPGTSLSTASLVTGPIAVQTTAGKQLSLTHPEILLVPOI-----S 459
Db 538 VTSEDTVYTTFCPLTETTPATESAPESAPATESVPATESAPVA---PESSAPETETAPA 594
Qy 460 TEGGISTERNRIVDA--TTGLIPLTSVPTSAKEMTKLGLVTAEYSPASRLSGTSPSPQTT 518
Db 595 TESAPATESPPVAPGTESSVPAPESSAPAT-----ESAPATESPPVAPGTETT 642
Qy 519 VVSTAED-----LAPKSATFAVOSSTQSP-----TTLSSSASVNSC----- 554
Db 643 PATPGAESTPVAPVAPESSAPAVESSVPVAPGVETTPVAPVAPSTAKTALSALVSTEGTIP 702
Qy 555 -----AVNPCLHNGECVADNTS 571
Db 703 TTLESVPATQPSANSSYTIASVSS 726
```


RESULT 14
 US-08-185-432-16
 ; Sequence 16, Application US/08185432
 ; Patent No. 5750652
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, Spyridon
 ; APPLICANT: Busseau, Isabelle
 ; APPLICANT: Diederich, Robert J.
 ; APPLICANT: Xu, Tian
 ; APPLICANT: Matsuno, Kenji
 ; TITLE OF INVENTION: DELTEx PROTEINS, NUCLEIC ACIDS, AND
 ; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PENNIE & EDMONDS
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/185,432
 ; FILING DATE: 21-JAN-1994
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Misrock, S. Leslie
 ; REGISTRATION NUMBER: 18,872
 ; REFERENCE/DOCKET NUMBER: 7326-006
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-8864/9741
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2471 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 US-08-185-432-16

[illegible]

```

Db      1060  CQTLVNLCSSRSPCKNCTCVQKKAESOCICLPSGWAGAYCDVPN--VSCDIAASRRGVLV 1111

RESULT 15
US-08-083-590A-19
; Sequence 19, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; PS-08-083-590A-19

```

```

Query Match      5.2%; Score 254; DB 1; Length 2471;
Best Local Similarity 23.7%; Pred. No. 1e-08;
Matches 85; Conservative 37; Mismatches 119; Indels 118; Gaps 14;

Qy 550 SVNCAVNPCLHNGECVADNTS----- 571
   : : || |||| : : :
Db 795 NIDECASNPCLNQGTCFDIDISGYTCHCVLPYTGKNCQTVLAPCSNPNCENAAVCKESPNF 854

Qy 572 RGYHCRPPSPHQDDCSVDVNECLSNPCPSTATCNNQGSFICKCPGYQ----- 621
   | | | | | | | | | | : : : : : : : : : : : : : : : : : :
Db 855 ESYTCLCAPGQGORCTIDIDICKPNHGLCHNTQSYMCPCPGFSGMDCCEEDIDD 914

Qy 622 -----LEKGTC-NLVRTF-----VTEPKLRTELNTTVEKHSDLQEVENETTKTLNMCFS 670
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 915 CLANPCONGSCMGVNTFSLCLPGF-----TGDKCQTDMECLSEPCCKNGTGCD 966

Qy 671 ALPYSIRST-----VHASRESNAVISLQTFSLASNVTLFDLADRMQKCVNSCKSAE 724
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 967 YVNSYTKCQAGFGVHCENNINEC-----TESSCFNGGT-----CVDGINFSC 1011

Qy 725 VCQLLGSQRRIFRAGSLCKRKRSPECDK-----DTSICTDLGDGVALCQCKSGYQFQNMKDH 780
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1012 LCP-----VGFTGSLCHEINECSSHPCLNEGTCVDGLGTYRCSPGLGTGN----- 1059

Qy 781 CRA-----CEDGYRLENETCMS-----CPFGLGLNCGNPYOLITVVIAGGGLLL 827
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 1060 CQTLVNLCRSRSPCKNKGTCYQKKAESQCLCPGSGWAGYCDVPN--VSCDIAASRRGVLV 1116

Search completed: July 23, 2002, 14:29:26
Job time: 140 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 14:27:56 ; Search time 25.95 Seconds
(without alignments)
3502.910 Million cell updates/sec

Title: US-09-840-746-1
Perfect score: 4859
Sequence: 1 MSQTEVRSVAPMRGEIT.....FPGQYNPSFISDESRDDYF 946

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	409	8.4	1104	2 S59310	probable membrane
2	399	8.2	1367	1 S48478	glucan 1,4-alpha-g
3	392.5	8.1	3507	2 T34513	hypothetical prote
4	387.5	8.0	2232	2 T34434	hypothetical prote
5	364	7.5	2271	2 F90073	hypothetical prote
6	357.5	7.4	1032	2 T34433	hypothetical prote
7	347	7.1	528	2 T47141	gastric mucin (clo
8	343	7.1	1161	2 S57180	probable membrane
9	342.5	7.0	573	2 A33533	cell surface glyco
10	339	7.0	786	2 T16509	hypothetical prote
11	328	6.8	1257	2 S28764	neurocan precursor
12	325	6.7	1306	2 S25370	MSB2 protein - yea
13	323	6.6	4776	2 E95206	cell wall surface
14	315.5	6.5	1268	2 S52781	neurocan - mouse
15	313.5	6.5	534	2 T39903	serine-rich protei
16	313	6.4	1802	2 S69703	HKR1 protein precu
17	312.5	6.4	1630	2 A53577	ascites sialoglyco
18	310.5	6.4	948	2 T11678	hypothetical prote
19	310.5	6.4	3570	2 T45025	mucin MUC5B, trach
20	309.5	6.4	1169	2 S38181	flocculation prote
21	309.5	6.4	3562	2 A47171	chondroitin sulfat
22	305	6.3	725	2 A41258	a-agglutinin core
23	303.5	6.2	1643	2 T14274	versican precursor
24	301	6.2	1367	2 S51959	hypothetical prote
25	300.5	6.2	1275	2 T33369	hypothetical prote
26	297	6.1	1459	2 T32271	hypothetical prote
27	295.5	6.1	610	2 PN0012	mucin 4, tracheal
28	293.5	6.0	796	2 T21460	hypothetical prote
29	293	6.0	1609	2 S25345	probable membrane

RESULT 1
S59310
probable membrane protein YMR317w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR924.09
C:Species: Saccharomyces cerevisiae
C:Date: 29-Nov-1995 #sequence_revision 23-Feb-1996 #text_change 05-Nov-1999
C:Accession: S59310
R:Churcher, C.M.
submitted to the EMBL Data Library, September 1995
A:Reference number: S59302
A:Accession: S59310
A:Molecule type: DNA
A:Residues: 1-1104 <CHU>
A:Cross-references: EMBL:254141; GSPDB:GN00013; MIPS:YMR317w
A:Experimental source: strain AB972
C:Genetics:
A:Gene: MIPS:YMR317w
A:Map position: 13R

ALIGNMENTS

RESULT 1

Query Match 8.4%; Score 409; DB 2; Length 1104;
Best Local Similarity 22.4%; Pred. No. 1.3e-11;
Matches 230; Conservative 160; Mismatches 385; Indels 254; Gaps 37;

Qy 1 MSQTEVRSVAPMRGGEITAHLLTNTTSADVTG-----SSASYPEG-VNASVLQ 52
Db 52 VSSSTLVSSVVP-----EFTSSSLSDTIASILSSESLVSIFFSLSYTSDISSTVND 107
Qy 53 FSDSTVQSGGSHLTALGDRSYSESSSTSSSESLNSSA-----PRGERSIAGISYGQVRGTA 107
Db 108 VESSTGSPSNYSALSSNTNAQLSSSTTTDTDISSSAIOITSSPQTSSNGGSSSEPLGKS 167
Qy 108 IEQRTSDHTHTVL-SSTFTKGERALLSITDNSSSDIVESSTSYIKISNSSHSEYSF 166
Db 168 SVLETTASSDSTTAVTSSTFT-----TLTDVSSSPKISSSGSAVTSVGTTSDAKKEVF 220
Qy 167 SHAQTERNISYDGEYQAPTESPVLHTSNLPSTPINMPNTSVLDTDAEFVSD--- 223
Db 221 SSSSTDVSSLLS-----STSPA-----SSTISETLPFSSTILSTSSPSSEAP 265
Qy 224 -----SSSSSSSSSSSSSGPLPLPSV-----SOSHLLFSILP-STRASVHL-LK 268
Db 266 SATSSSSSEASSSTSSSVSEAPLATSSVSSEAPSTSSVSSEAPSSSSSSSSSEIS 325
Qy 269 STSDASTPWSSPPLVSLTSTSAPLSVSQT---LPQSSSTPV---LPARETPVTS 322
Db 326 STTSSV---SSEAPLATSSVVSSEAPSTSSVSSEAPSTSSVSSEAPSSSTSSSVSS 382
Qy 323 FQTSTMTSFTMLHSSQTADLKSOQSTPHQKVITEKSPSLVSLPTESTKAVTNSPLPP 382
Db 383 EISSTKSSVMSSEVSSATSSSLVSSSEAPSAISSLASSRLFFSSKNTSVTSTLVATEASSVTS 442

RESULT 4
T34434
hypothetical protein K06A9_la - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: T34434
A:Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA

RESULT 4
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-oct-1999 #te
C:Accession: T34434
R:Geiselt, C.; Gattung, S.
A:Description: The sequence of C. elegans cosmid K06A9
A:Reference number: Z21525
A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA

A; Map position: X
A: Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2; 99

Seq. Loc. Similarity	Seq. NO.	Seq. NO.
Matches 207; Conservative 143; Mismatches 352;	Indels 221; Gaps 34;	

Qy	6	TVSRVAPMRGEITAHWLTJNST-----SADVTGSSASYPEGVNASVLTFQSSTVQSG	60
Db	246	SISTSALPIASSSSSPSAASSTTPVLSSTIQSSGTFPSSVASSPSTVGSTVSGAAS	305
Qy	61	GGSHITALDGRSYSESSST-----SSSLNSASPRGRSTAGISYGOVGRATEQRTSSDH	116
Db	306	SSSYATVSTIAGSTGTTTPVGGSSSTIGSSTPSASSSSGTM-----STISGGTGSTV	359
Qy	117	TDHTYLSSTTKGERALLSITDINSSSDIVEST--SYIKINSSSHSEYSSPSHAQTERS	174
Db	360	TVPGCSSTF-----ASSTPIASSSPGSTVTVAPGSSSTYGSSSTPSSASSSS	406
Qy	175	N--ISSYDGEYAQPSTESPVLHNTSNLPYPTINMPNTSVLTDAEFYSDSSSSSSSSS	232
Db	407	SGTMSTNGSGTGSTVTVAPV--SSSTFGSSTPIASSSSSG-----STVVVVGSSSTVGSST	461
Qy	233	SSSSSGPPLPLPVSQSHHLFSSILPTSTRASVHLLKSTSDASTPWSSSPSPVSLTTST	292
Db	462	PSASSSSAGTASTISGSGTATVPGSSSV-----GSGTQASPS--PGTMST-V	511
Qy	293	SAPLUSVQOTTLPOGSSTPV--LPRARETPVTGFTSTMTSFMTMLHSSQATADLKSQSTP	349
Db	512	SGPTGSTVTVVVGSSSTGPAPSSPNPSSSPASTGSTITTSGSSSIIVTVVSGSTVSGSG	571

QY	350	HOEKVITESKSP--SLVSLPTESTKAVTNTNPLP--PSLTSTSTSTQTLPLATSTNLQAQSP	405
DB	572	TSQSTLASSTATPGSSSTVPSSSQSPAPNTGTTPSOTSSQSPSPMNPSSSTP	631
QY	406	TFTTILKTSQPLMTTPGTLSTASTSLVVG---PIAVQITAGKQLSLTHPEILYPOISTEG	462
DB	632	T-----GSSOSTIPEGSTASSPGSTGTSFSAATEVTS-----QSTVPSSGLG	676
QY	463	GISTERN---RVIVDATGLTGLTSPVTSKAKEMTTKLGVTAEYSPASRSLSLGTSPSPQTTV	519
DB	677	TOSTNSPSPSSLSFSTGMSLTLTSEPS-----SSTQSSGAQSTLTTPSPN---	723
QY	520	VSTAEDLAPKATFAVQSGSTQSTPTLSSASVNSCAVNPCLHNGECVADNTRSGYHCRCP	579
DB	724	-----PSQSTSSLESSTGATTSAGSAGTTWTS-----	751
QY	580	PSWGGDDGSDVNECLSNPCPSTACNNT--QGSFTCKCPGVQLEKGCICNLVTRTVTEFK	638
DB	752	PS---QSSSVSSQSGSTSPAASTTSGEMTSQGS-----TOTPCSSVS	790
QY	639	LKRTPLNTVEKHSDQLQVENEITKLNMCFSALPSYIRSTVHASRESNAVVI-----	691
DB	791	TSAAILTSTQQSVS--TNSPGSTVTRP-----SIVSGSTSGSTVTVGSGTEAS	836
QY	692	-----SLQTTFSLASNWTLPDLADRQK-----CVNSCKSSAEVCCOLLGSRRI-----	735
DB	837	TSGSLATTAPKPSVTCLFMYDTQSKIEDQTAINTYKTFNFALLVASKLNNESSILGY	896
QY	736	-----FRAGSLCKRKRSPECD-----KDTSICTDLGDGVALCOCKSGYQFNKM	777
DB	897	IDNFGYSAGLNDHQYYPTDDYNGIKSVFPFDGTDGDDIDLQKDV-----	941
QY	778	DHSCRACEDGYRLENTCM-----SCPFG--LGLGNCNPNYQLITVVIAAAGGGLLILGIA	832
DB	942	DKSLATADWTPPVADOTCMFISARPEDEYGGTTIKSTYTYFETVGVVLGGAKSIPGLS	1001
QY	833	----LIVTCCRKNKNNDISKLIFFK	851
DB	1002	IDKNIVITNTMTDRDASAVVSK	1024
RESULT	7		
147141		gastric mucin (clone PGM-2A) - pig (fragment)	
C:Species:	Sus scrofa domestica (domestic pig)		
C:Date:	21-Feb-1997	#sequence_revision	21-Feb-1997 #text_change 03-Nov-2000
C:Accession:	I47141; S55315		
R:Turner,	B.S.; Bhaskar,	K.R.;	Hadzopoulou-Cladaras, M.;
Gastroenterology	106,	200,	1994
A:Title:	Pig gastric mucin: isolation and characterization of a cDNA clone with a novel		
A:Reference number:	I47141; MUID:94102478		
A:Accession:	I47141		
A>Status:	preliminary; translated from GB/EMBL/DBJ		
A:Molecule type:	mRNA		
A:Residues:	1-528	<TUR>	
A:Cross-references:	EMBL:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208		
R:Turner,	B.S.; Bhaskar,	K.R.;	Hadzopoulou-Cladaras, M.;
Biochem. J.	308,	89-96,	1995
A:Title:	Isolation and characterization of cDNA clones encoding pig gastric mucin.		
A:Reference number:	S55315; MUID:95275264		
A:Accession:	S55315		
A>Status:	preliminary		
A:Molecule type:	mRNA		
A:Residues:	1-528	<TU2>	
A:Cross-references:	GB:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208		
C:Superfamily:	pig submaxillary mucin		
Query Match		7.1%	Score 347; DB 2; Length 528;
Best Local Similarity		27.9%	Pred. No. 4e-09;
Matches 161; Conservative		81; Mismatches	204; Indels 132; Gaps 23;
QY	10	SVAPMRGGEITAHWLLTNTSTTSADVTGSSASYPGCVNASVLTQFSDS-----TVQSGG	62

Query Match . 7.1%; Score 347; DB 2; Length 528;
Best Local Similarity 27.9%; Pred. No. 4e-09;
Matches 161; Conservative 81; Mismatches 204; Indels 1

QY 10 SVAPMRGGEITAHWLLTNSTTSADVTGSSASYPEGVNASVLTTQFSDS-----TVQSGG 62

Db 35 SVQPSSSGSA-----TTSATSVQ-TSSSSPPISSITISVQTSSSSVPTTSTTSVQPS 88
Qy 63 SHTALGRSVSESSSTSSSLNAPSAGERSIAGISGVQVGRGTAEIQRSSDHTDHYL 122
Db 89 SSSAPTRATSVQSSSSSAPISSTT-----SVQPSGGSVPTTTSATSVQSSSSSAPT 143
Qy 123 SSTFTKGERALLSTIDNSSDDIVESSTYKIKSNHSHEYSFSSHAQTERSNISSYDGE 182
Db 144 SAT-----SVQPSSSSPPISSIVQPSSSSAPTTSATS----- 179
Qy 183 YAOPTSES--PVLHNSLNPYPTINMPNTSVVLDTDAEFVSDSSSS-----SSSSSSSS 236
Db 180 -VQPSSSSSPPI-----SSTVSQTSSSSVPTTTSVQPSSSSVPTTSATSVKSS 231
Qy 237 SGPPPLPVSQSHHLLFSSLPSTRASVHLLKTSDDASTPSSSPPLP-----VSLTST 292
Db 232 SSSSTPIPT-----TSVQPSSSSAPTTSATS--VQPSSSSSTPIPTTSVQPSSS 282
Qy 293 SAPLSVQTTLPQSSSTPVLPRARETPTVSFOTSTMTSFMTMLHSSQTADLKOSTPHQE 352
Db 283 SAPITSATSVPSSSSP-----PISS-----TISVQPSSSSSSP 317
Qy 353 KVITESKPSLV-SLPTESTKAVTNSPLPPLPSLTESSTEQTLPATSNLAQMSPTFTTTI 411
Db 318 TTSTTSVQPSSSGAPTTSATSVPSSSSSPPI--SSTISVQPSSSS-----SPTTSTTS 371
Qy 412 LKTSQPLMTPTGTLSSPASVLVGTGPIAVTGTAGKQLSLTHPEILVPOISTEGGISTERNRV 471
Db 372 VQPS-----SSGSAPTTSATSVPSSSS-----VPTTSATSVKSSSSST 412
Qy 472 IVDATTGLIP--LTSVPT-----SAKEMTKLGVTAEYSPASRSLSGTSPSPQTTVVST 522
Db 413 PIPTTSVQPSSSSVPTTSATSATSQTSSTSTPIPTTSVQPSSSSAPTTS-ATSVPQS 471
Qy 523 AEDLAPKSAFAVQ--SSTQSPITLSSASVNSCAVNP 558
Db 472 SSSSPPISSITISVQPSSSSSSPTTTSVQPSSSGSAP 509
RESULT 8
S57180
probable membrane protein YJR151c - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein J2233; serine/threonine-rich protein YJR151c
C;Species: Saccharomyces cerevisiae
C;Date: 23-Aug-1995 #sequence_revision 08-Sep-1995 #text_change 05-Nov-1999
C;Accession: S57180
R;Scarcez, T.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S57169
A;Accession: S57180
A;Molecule type: DNA
A;Residues: 1-1161 <SCA>
A;Cross-references: EMBL:Z49651; NID:g1015902; PID:g1015903; GSPDB:GN00010; MIPS:YJR151c
C;Genetics:
A;Gene: MIPS:YJR151c
A;Map position: 10R
C;Keywords: transmembrane protein

Query Match 7.1%; Score 343; DB 2; Length 1161;
Best Local Similarity 21.3%; Pred. No. 1.6e-08;
Matches 204; Conservative 147; Mismatches 392; Indels 216; Gaps 34;
Qy 3 QTEVRSVAP-----MRGGEIT-----AHWLLNNTTSADVTGSSASYPE 43
Db 60 KTEYPSIEAAAVFDYGDFTRLTGISGDEVTRMITGVPMYSTRLKPAT-----SSALSKD 115
Qy 44 GVNASVLTFQSDSTVQSGGSHALGCD-RSYSESSSTSSSLNAPSAPRGERSIAGISYGQ 102
Db 116 GIYTAIPTSTSTTTKSTSTPTTTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 175
Qy 103 VRGTAIEQRTSSDHTDHT-----YLSSTFTKGERALLSIDNSSSDIVESSSTYKISN 157

Db 176 TSTPTTSTSTPTTSTPTTSTPTTSTPTTSTPTTSTPTTSTPTTSTPTTSTPTTSTPTTST 235
Qy 158 SSHSEYSSFSHAQTERSNISYDGEYAQPSTE----SPVLHNSLNPSTPTPIN-MPNTSV 212
Db 236 TSTTSQSTKSTPTTSTPTTSTPTTSTPTTSTPTTSTPTTSTPTTSTPTTSTPTTSTPTTST 295
Qy 213 VLDTDAEFVSDSSSSSS-----SSSSSSSGPPPLPVSQSHHLLFSSILPSTRASVHLL 267
Db 296 ---TSSFTSSASASSSVISTTATSTTFASLTTPATSTASTDHTTSSV--STTNAFTTS 350
Qy 268 KSTSDASTPSSSPPLPVLSTTSAPLSVQST--LQOSSSTPVLPRARETPTVSFOT 325
Db 351 ATTTTSDTVIISSSP--SQVTSSAEPTTVSEVTSVVEPTRSSQVTSABEPTTVSEFTS 407
Qy 326 STMTSFMTMLHSS---QTADLKQSTPHQEKVITEKSPSLVS-----LPTSESTKAVT 375
Db 408 SVEPTRSSQVTSAAEPTTVSEFTSSVEPTRSSQVTSAAEPTTVSEFTSSVEPTRSSQVTS 467
Qy 376 TNSPLPPLSTESSSTEQTLPATSTNLAQMSPTFTTTILKTSQPLMTPTGTLT---STASLV 432
Db 468 SAEPITVSEFTSSVEPTRSSQVTSAAE--PTTVSEFTSSVEPTRSSQVTSAAEPTTVSEV 525
Qy 433 TGPIAVOTTAGKQUSLTHP-----EILVPOISTEGGISTERNRVIVDA-- 475
Db 526 TS--SVEPIRSSQVTTTTEPVSSFGSTFSEITSSAEPLSFASKATTSAESISSNQITISSEL 583
Qy 476 -TTGLIPLTSTVPTSAKEMTKLGVTAEYSPASRSLSGTSPSPQTTVVSTAEADLAPKSAF 534
Db 584 IVSVITSSSEIPSEIEVLTSGLSSSVSEPTSL---VGPSSDESISTESLSATSTFSA 640
Qy 535 VQSS-----TQSPITLSSASVNSCAVN-----PCLHNGECVADNTSRGYHCRC 578
Db 641 VVSSSKAADFTTRTVSAKSDVNSSTTQSTTFEATPSTPLAVSSTVVTSST----- 692
Qy 579 PPSWQGGDDCDV--NECLSNPCSTATCNTNQCSFTCKCPVQLEKGICNLVRT---F 633
Db 693 -----DSVFNIPFSEISSPESSTA-ITSTSTSTAE-----RTSLSY 730
Qy 634 VTBEFKRTFLNTTVEKHSDLOEVEINEITKLNKCFASALPSYI--RSTVHASRESNAVVI 691
Db 731 LSSNMSSTFTLSTFTVSQSVSSFSMEPTSSVASFSSPPLLVSRNCSNARSNTISS 790
Qy 692 SLQTTFLASNVLT-LFDLADRMQKVNCKSAEVCOLLGSSQRRIFRAGSLCKRKSPECD 750
Db 791 GLRSTIENVRNATSTFTNLSTDEIVITCKSS-----CT 824
Qy 751 KDTISCTDLDGVALCQCKSGYFQFNKMDHSCRACEDGY----- 788
Db 825 NEDSVLTKT-----QVSTVETITSCSGGICTTLMSPVTTINAKANTLTTE 871
Qy 789 -RLENETCMSCPPGLGNGPNPYQLITVIAAAGGGLLLILGIALIVTCCRKKNKNDIS 846
Db 872 TSTVETITTCPGGV-----CSTLTVPVTTITSEA-----TTTATISCEDNEEDIT 917
RESULT 9
A33533
cell surface glycoprotein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 29-Jan-1990 #sequence_revision 29-Jan-1990 #text_change 11-Jan-2000
C;Accession: A33533
R;Dougherty, G.J.; Kay, R.J.; Humphries, R.K.
J. Biol. Chem. 264, 6509-6514, 1989
A;Title: Molecular cloning of 114/A10, a cell surface antigen containing highly conse
lines.
A;Reference number: A33533; MUID:89197960
A;Accession: A33533
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-573 <DOU>
C;Cross-references: GB:J04634; NID:g191943; PIDN:AAA37239.1; PID:g309106
C;Superfamily: unassigned EGF-related proteins; EGF homology

C; Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C; Accession: T16509
R; Nhan, M.
submitted to the EMBL Data Library, December 1995
A; Description: The sequence of C. elegans cosmid F59A6.
A; Reference number: Z18526
A; Accession: T16509
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 1-786 <NHA>
A; Cross-references: EMBL; U01994; NID: g1123047; PID: g1123052; PIDN: AAA03456.1; CESP: F59A6
C; Genetics:

A:Residues: 1-1257 <RAW>
A:Cross-references: EMBL:M97161; NID:g205649; PIDN:AAC37679.1; PID:g205650
C:Superfamily: aggrcan; C-type lectin homology; complement factor H repeat homology; EQ
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-1257/Product: neurocan #status predicted <MAT>
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-355/Domain: link protein repeat homology <LNK2>
F:364-366/Region: cell attachment (R-G-D) motif
F:953-984/Domain: EGF homology <EGF>
F:1029-1149/Domain: C-type lectin homology <LCH>
F:1156-1212/Domain: complement factor H repeat homology <FHD>
F:121339/737,967,1164/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:372,410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F:944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

Query Match 6.8%; Score 328; DB 2; Length 1257;
Best Local Similarity 22.0%; Pred. No. 8.6e-08;
Matches 155; Conservative 107; Mismatches 255; Indels 186; Gaps 31;

QY 1 MSQETVRSVAPMRGEITAHMLLTNSTTSADVTGSSA-----SYPEGVNASVLTFQF 53
DB 418 LTRTQASQETLASTPGPTLASMLLTGVTSSTGVPSPSSLCVDMETTPSG-----TQV 471

QY 54 SDTVOGGGHTALGDRSYSESSSTSSSLNSSAPRGERSIAGISYGVQVRGTAIEQRTS 113
DB 472 APTPTMRGRFGKGLNGRHFGQOG--PEDQLLEAAEASQAQPTTEVTAHMGPSAATEALE 529

QY 114 SDHTDHTYLSSTFKGERALLSTDNSSSDIVESSTSYIKINSSHSEYSSFSHAQTER 173
DB 530 SDQSHSPW-----AIL-----TNEVDVPCAGS----LGRSLPESRKWSPLISP 570

QY 174 SNISSYD-----GEVAQSPSTESPVLTHTSNLPSYPTINMPNTSVVLDDAEFVSDSS 226
DB 571 STVPSTDTPLKPGDAEPKVASATHPPWLPFS-EPAV-----PSSIPSEALS 618

QY 227 SSSSSSSSSSSGGPLPLPSYSQSHILF----SSILPSTRASVHLKSTSDASTPWSSSPS 282
DB 619 AVSLQASPGGDPDFPIVAMLRAPKLWLLPHSTLVPNV-----S 657

QY 283 PLPVSLTTSAPLSVSQITLPOSSSTPVLPRARETPVTSFQSTWTFMTLHLSQATD 342
DB 658 PIPLS-----PASPLPSSVPEEQAVRPVFGAEDPETPQOT-TMAAPGEASHGSPED 709

QY 343 -----LKQSQTPHOEKVITESKPSLSLPTES--TKAVTNSPLPPSLTESSTEQ-- 391
DB 710 SIEIGISSQATKHP-----ISGPNASLDSSNVTVN-VPVSDAGILGTESGV 756

QY 392 -TLPATSTNLQAQSPFTTTILKTSQPLMTTGGTSLSTASLVTGPIAVQTAGKQLSLTH 450
DB 757 LDLPGSPSTDGQAT---VDMVLATWLP---PCHGLDTGSQST-PMEAH---GVTKSVE- 805

QY 451 PELVLPQISTEGIGISTERNRVIIVDATTGLIPLTSVPTSAKE-----WTKLGVTA---EY 502
DB 806 -----PTVALEGATKDP-----MEATMDVVPSTVDATISGSEPKSSISTHHVVTAAGDQG 856

QY 503 SP-----ASRLGTSPTS---POTTVVSTAEDLAPKSAATFAVO---SSTQSP 542
DB 857 TPTLTPTSSEGVVAQESLGLTSLSPHPWSSSLASSDEAVASVSSGEPFLWDIPSTLIP 916

QY 543 TLLSSAS-----VNSCAVNPCLHNGECVADNTRSGYHCR 577
DB 917 VSLGLDESOLKVVVAESPGLEGEVAVASQGEDPTDCENNPNCLHGTCRTNGTM--YGCS 974

QY 578 CPPSWGDDCSVDVNECLSNPCPSTATCNCNTQSFICKCPGVY 620
DB 975 CDQGYAGENCEIDIDCLCSPCENGCTCIDEVNGFICLCLPSY 1017

RESULT 12
S23370
MSB2 protein - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein G4017; protein YGR014W
C:Species: Saccharomyces cerevisiae
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000
C:Accession: S25370; S64305
R:Bender, A.; Pringle, J.R.
Yeast 8, 315-323, 1992
A:Title: A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence defect.
A:Reference number: S25370; MUID:92383951
A:Accession: S25370
A:Molecule type: DNA
A:Residues: 1-1306 <BEN>
A:Cross-references: GB:M77354; NID:g171993; PIDN:AAA34798.1; PID:g171994
R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64071
A:Accession: S64305
A:Molecule type: DNA
A:Residues: 1-1306 <BEN>
A:Cross-references: EMBL:272799; NID:g1322977; PID:g1322978; MIPS:YGR014W
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:MSB2
A:Cross-references: SGD:S0003246; MIPS:YGR014W
A:Map position: 7R
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida
C:Keywords: transmembrane protein
F:3-19/Domain: transmembrane #status predicted <TM1>
F:1189-1205/Domain: transmembrane #status predicted <TM2>

Query Match 6.7%; Score 325; DB 2; Length 1306;
Best Local Similarity 27.2%; Pred. No. 1.2e-07;
Matches 151; Conservative 95; Mismatches 221; Indels 88; Gaps 21;

QY 28 STTSADVTGSSASYPEGVNASVLTFQSDSTVQSGGSH-TALGDRSYSESSSTSSSESLNS 86
DB 494 SQSSSDVATSA---PSVWSSS---FSYTLQAGGSSMTNPSSSTIYVSSSTGSSE--ES 545

QY 87 SAPRGERSIAGISYGVQVRGTAIEQRTSSDHTHTLSSTFKGERALLSTDNSSSDIV 146
DB 546 AASTASATLSGSSSTYMAGNLQSQPPTS-----SLLSEQATSTSAVL 589

QY 147 ESTSYIKINSHSEYSSFSHAQTERSN-ISSYDGYAQP--STESPVLTHTSNLPSYTP 203
DB 590 ASS-----SVSTTSPYTTAGGASTEASSLSSTSAETSQVSVSQSTTALQTSFASST 643

QY 204 TINMPNTSVVLDDAEFVSDSSSSSSSSSSSSSGPPLPLPSVYSQSHLPSILPSTRAS 263
DB 644 TEGSETSSQGFSTSSVVLQMPSSLSSEFSPQTQMNASASSSQ-----YTISS 693

QY 264 VHLKSTSDASTPWSSSPPL-----PVSLTTSTAPLSVSTQTLTLPQSSSTPVLPRARE 317
DB 694 TGIQSVDSTSVSYTSSSSSVQSDTPVSYTSSSSSVQSDTPVSYTSSSSSVQSD 753

QY 318 TPTVTSQTS-----TMTSEFMTLHS-SQTADLKSQSTPHQEKVITESKPSLSVLP 367
DB 754 TPV-SYTTSSSSSVQSDTPVSYTSSSSSVQSDTSPVSTSSRSSSVQSDTP-----VP 808

QY 368 TESTKAVT--TNSPLPPLSTES-----STEQTLPATSTNLQAQSPFTTTILKTS---QP 417
DB 809 STSRSSVSQTSLSLQPTTSSQRTTSTHGAL-SESSSVSQQASEITSSINATASYHS 867

QY 418 LMTTPGTLST-----ASLVTGPIAVQTAGKQLSLTHPEILVPOISTEGGISTERN- 469
DB 868 IQTTAATQSTTLSTFDANSSASAPLEAVATSTPTPSKASALLTPTSTSLSQVATNV 927

QY 470 --RVIVDATTGLIPLTSVPTSAKEMTKLGVTAEYSPASRSLCTSPQSTPVVSTABDLA 527
DB 928 QTSLLTTESTTVLEPSTINSSSTESLVT--SSDNNMWIPTELITOAPEAASSTASVTGGT 985

QY 528 PKSATFAVQSSQSP 542
DB 986 TMTLPHAIANAATQVP 1000

RESULT 13

cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain TIGR:SP4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: E95206
R: Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidelberg, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, M.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Fitts: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: E95206
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4776 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75846.1; PID:g14973269; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1772

Query Match 6.6%; Score 323; DB 2; Length 4776;
Best Local Similarity 24.0%; Pred. No. 7.2e-07;
Matches 137; Conservative 108; Mismatches 285; Indels 40; Gaps 9;
QY 1 MSQETVSRVSAPMRGGEITAHLWLLTNTSTTSADVGTSSASYPGCVNASVLTFQSDSVQS 60
DB 2691 ISASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASTSASES---A 2747
QY 61 GGSHALGDRSYSESSSTSSSELNSAPRGERSIAGISYQVGRGTAIEORTSSDTHDT 120
DB 2748 STSTASASTSASESASTSASASASTSASASTSASASTSASASTSASASTSASESAST 2803
QY 121 YLSFTFKGERALLSIITDNSSSDIVESSTSYIKINSNSHSEYSSFSHAOTERSNISSYD 180
DB 2804 SASASASTSASASASTSASASASTSASASASTSASASTSASASTSASASTSASESAST 2859
QY 181 GEYAPQSTSPVLHTSNLPSYTPPTINNPNTSVVLDTDAAEFVSQSSSSSSSSSSSSSSGPP 240
DB 2860 SASASASTSASESASTSASESASTSASESASTS-----ASASASTSASASTSASA 2909
QY 241 LPLPSVQSQHLLFSSILPSTRASVHLKLTSDASTPWSSSPPLPVLSTSTGAPLSVQ 300
DB 2910 SASTSASESASTSASASTSASASTSASESASTSASESASTSASESASTSASESASTSASES- 2968
QY 301 TTLPOSSSTPVLPRARETPVTSTQTSMTSMTMLH-----SSQADLKSSQSTPHQ 351
DB 2969 TSASESASTSASASTSASESASTSASESASTSASESASTSASESASTSASESASTSASESASTSAS 3028
QY 352 EKVITE-SKSPSLVSLPSTESKA-----VTTNSPLPPLPSTESSTEQPLPATSTNLAQMS 405
DB 3029 ASASTSASASASTSASESASTSASESASTSASESASTSASESASTSASESASTSASESASTSASES 3088
QY 406 TPTTTLKTSOPLMTTPTGTLSSSTASVLTGPIAVOTTAGKOLSLTHPILPQISTEGGIS 465
DB 3089 TSASASTSASESASTSASESASTSASESASTSASESASTSASESASTSASESASTSASESASTSAS 3148
QY 466 TERNRVIVDATTGLPLTSVPTSAKEMTKLGVTAESVSPASRLGTSPTQTTVVSTAE 525
DB 3149 ASTS---ASASASTSASESASTSASESASTSASESASTSASESASTSASESASTSASESASTSASA 3205
QY 526 LAPKSAFAVOSTQSPPTLLSSASVNSCA 555
DB 3206 STSASASTSASESASTSASESASTSASESASTSASESASTSASESASTSASESASTSASESASTSASA 3235

RESULT 14

S52781
neurocan - mouse

C:Species: Mus musculus (house mouse)
C:Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 04-Feb-2000
C:Accession: S52781
R: Rauch, U.; Forsberg, N.; Kulbe, G.; Arnold-Ammer, I.; Faessler, R.
submitted to the EMBL Data Library, February 1995
A:Description: Amino acid sequence of mouse neurocan and brevican and their different
A:Reference number: S52781
A:Accession: S52781
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1268 <RAU>
A:Cross-references: EMBL:X84727; NID:g758629; PIDN:CAA59216.1; PID:g758630
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-355/Domain: link protein repeat homology <LNK2>
F:964-995/Domain: EGF homology <EGF>
F:1040-1160/Domain: C-type lectin homology <LCH>
F:1167-1223/Domain: complement factor H repeat homology <FHD>

Query Match 6.5%; Score 315.5; DB 2; Length 1268;
Best Local Similarity 23.0%; Pred. No. 3.3e-07;
Matches 164; Conservative 95; Mismatches 259; Indels 195; Gaps 31;
QY 1 MSQETVSRVSAPMRGGEITA-----HWLLTNTSTTSADVGTSS-----ASYPEGVNA 47
DB 418 LTWTQAPETLTGTPGPTTLASWPSSSEKWLFTGAPSMGVSSPSDMGVDMETPLG--- 474
QY 48 SVLTQFSDSTVQSGGSHALGDRSYSESSSTSSSELNSAPRGERSIAGISYQVGRGTA 107
DB 475 ---TQVAPTFTMRGRFKGLNGRHFQOQGPEDQLPEV--AEPSAQPTTLGATANHMRPSA 529
QY 108 IEORTSSDHDTHYLSFTFKGERALLSIITDNSSSDIVESSTSYIKINSNSHSEYSSFS 167
DB 530 ATEASESDQSHSPWAILTNEVDFGAGSLGRSLPESLMSPS---LISPVPS----- 580
QY 168 HAOTERNISYDGEYAQAPSTESPLHTSNLPSYTPPTINNPNTSVVLDTDAAEFVSQSSSS 227
DB 581 ---TE-STPKFGAAPSVKSAIPLPLPPEPPA---PSPG-----PSEALSA 624
QY 228 SSSSSSSSSGPPPLPLPSVQSQHLLF-----SSILPSTRASVHLKLTSDASTPWSSSP-S 282
DB 625 VSLQASSADGSPDFIVAMLRAPKLWLLPRSTLVN-----MTPVPLSPAS 670
QY 283 PL-----PVSL-----TTSTAPLSVSQTLTPOSSSTPV----LPRARET 318
DB 671 PLPSWVPEQAVRPSVSLGAEDLETPFTTAAAPVEASHRS-PDADSTEIEGTSSMRATKH 729
QY 319 PVTSTFQSTWTSFMTM-----LHSSQTA----- 341
DB 730 PISGPWASLDSSNVMTMNPVPSDAGILGTESGVLDLPGSTSGGQATVEKVLATWLPPLGQ 789
QY 342 --DLKSQSTPHQEKVITEKSPSLVSL-----PTSTKAVTNTNSPLPSTLESSTEQ 392
DB 790 GLDPGQSQTPEAHGVAVSMET--VALEGGATEGPMETREV-----PSTADATWE-- 840
QY 393 LPATSTNLAQMSPTFTTILKTSOPLMTTPTGTLSSSTASVLTGPIAVOTTAGKOLSL-THP 451
DB 841 ----SESRSAISSTHIAVTMARAGCMPTLTSTSEGHPEPKQMVQAESLEPLNTLPSHP 896
QY 452 --ETLVPOISTEGGISTERNRVIVDATTGL--IPLTSVPTSAKEMTKLGVTAESVSPASR 507
DB 897 WSSLVVPDMEVASVSSE-----PTGLWDIPSTLIPVSLGLDSEVLNVAAE----- 942
QY 508 SLGTSFSPQTTVVSTAEADLAPKSAFAVOSTQSPPTLLSSASVNSCAVNPCLHNGECVA 567
DB 943 ----SPS----VEGFWEVA-----SQQEDPT-----DPCENNPCLGGTCHT 977
QY 568 DNTSRGYHCRCPSPSWQGDGCSVDVNECLSNPCPSTATFCNNTNQSGFICKCPVGY 620
DB 978 NGTV--YGCSCDQGYAGENCEIDIDCLCSPCENGCTCIDEVNGFICLCPLPSY 1028

RESULT 15

T39903
serine-rich protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000
C:Accession: T39903
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, November 1998
A:Reference number: 221889
A:Accession: T39903
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-534 <LYN>
A:Cross-references: EMBL:AL035354; PIDN:CAA22127.1; GSPDB:GN00067; SPDB:SPBC215.13
A:Experimental source: strain 972h-; cosmid c215
C:Genetics:
A:Gene: SPDB:SPBC215.13
A:Map position: 2
C:Superfamily: pig submaxillary mucin

		Query Match	6.5%; Score 313.5; DB 2; Length 534;
		Best Local Similarity	28.3%; Pred. No. 1.4e-07;
		Matches 146; Conservative	90; Mismatches 107; Indels 93; Gaps 19;
Qy	59	OSGSGHTALCDRYSSESSTSSSELNASSAPRGERSIAGISYGVRGTAEQRFSDDHDTD 118 :: : : : : : :	
Db	65	ETGGTPTYYG---YATPTSSEPFIFSESATPSETN----SYSPVSVSYSDPATSQLPSS 117 : : : : : : : :	
Qy	119	HTYLSSST---FTKGERALLSTIDNSS--SSDIVESTSIIYIKISNSHSYSFSHAQTORS 174 : : : : : : : :	
Db	118	TSFPSPTSSSEYTPSTESSLLDPSSVSAILPESTS-VEVISLSSSLSSDDPLTSSTFS 176 : : : : : : : :	
Qy	175	NISSYDGEYAQPSTESPVHLTHSNLPSTPTINMPNTSVVLTDAAEFVSDSSSSSSSSSSS 234 : : : : : : : :	
Db	177	SLSSTSS-SQPSVSS--TSSTTFSSAAPT-----STSSYSLSSSVWSSSSSPS 223 : : : : : : : :	
Qy	235	SSSGPPLPLPSVQSQHILFSSILSTRAS----VHLLKSTDASTPNSSSPPLPVSLTT 290 : : : : : : : :	
Db	224	SSSSTLTSSGLS-----TSSIPTSSSSSSTSSLSLSSSSSTACSSSSSSSIISSSS 277 : : : : : : : :	
Qy	291	STSAPLVSOTLTPQSSSTPVLPRARPTPTVTSFOTMTSMTHLHSSQTADLKXSQSTPH 350 : : : : : : : :	
Db	278	SSSSSPSTSTSTISSSSSSSSPSTSTSTTSS-SSSSSSSFSSVTLSSS----- 324 : : : : : : : :	
Qy	351	OEKVITSKSPULSVLPTESTKAVTNPSLPSPSTESTEQTLPATSNLAQMSPFTTT 410 : : : : : : : :	
Db	325	-----SMSSSSSSFSSPTSSSTTSSSSSPSSSFSSST-----TSSSKSSSPSST 371 : : : : : : : :	
Qy	411	ILKTSQPLMITPGTLLSSTASLVTCPIAQVTAGKQLSLTHPEILLVPQISTEGGISTE--- 467 : : : : : : : :	
Db	372	VSSSSS---TSSSTLTSSSSSSSRPAS-----SSSHSSSLSSHHKSSSSSSKSSAPV 419 : : : : : : : :	
Qy	468	RNRVIDVATGLIPTGVPTSAKEWTMLKGVTAEYSPASRLSGTSPS----PQTTVVSTA 523 : : : : : : : :	
Db	420	SSAFYHNSTSRSSSHSSHLSLSSKPILTA---SSSLLTSSSHITYERSTVVVVTV 475 : : : : : : : :	
Qy	524	EDLAPKGAATFAVGQST-----QSPRTLSSBAS 550 : : : : : : : :	
Db	476	ETVSSGSSTVASQSTQTSILLIVIGDSSSTDSGAS 511 : : : : : : : :	

Search completed: July 23, 2002, 14:30:42
Job time: 166 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 14:28:52 ; Search time 15.08 Seconds

(without alignments)
2428.959 Million cell updates/sec

Title: US-09-840-746-1

Perfect score: 4859

Sequence: 1 MSQTETSVKSVAPMRGGEIT.....PPQYNPFSFISDRRRDYF 946

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	399	8.2	1367	1 AMYH_YEAST	P08640 saccharomyc
2	396	8.1	1140	1 YMG6_YEAST	Q04893 saccharomyc
3	343	7.1	1161	1 DANA_YEAST	P47179 saccharomyc
4	342.5	7.0	573	1 C114_MOUSE	P19467 mus musculu
5	328	6.8	1257	1 PGCN_RAT	P55067 rattus norv
6	325	6.7	1306	1 MSB2_YEAST	P32334 saccharomyc
7	315.5	6.5	1288	1 PGCN_MOUSE	P55066 mus musculu
8	315	6.5	1802	1 HKR1_YEAST	P41809 saccharomyc
9	309.5	6.4	1169	1 YK82_YEAST	P36170 saccharomyc
10	309.5	6.4	3562	1 PGCV_CHICK	Q09053 gallus gall
11	305	6.3	725	1 AGAL_YEAST	P32323 saccharomyc
12	295.5	6.1	610	1 MUC4_HUMAN	Q99102 homo sapien
13	293.5	6.0	796	1 YS8A_CAEEL	Q09625 caenorhabdi
14	293	6.0	1609	1 FIG2_YEAST	P25653 saccharomyc
15	293	6.0	3381	1 PGCV_BOVIN	P81282 bos taurus
16	291.5	6.0	797	1 VGLX_HSVB	P28968 equine herp
17	285	5.9	1419	1 ALAL_CANAL	Q13368 candida alb
18	284.5	5.9	1537	1 FLO1_YEAST	P32768 saccharomyc
19	282.5	5.8	1322	1 YAG3_YEAST	P39712 saccharomyc
20	281.5	5.8	1260	1 ALS1_CANAL	P46590 candida alb
21	279	5.7	2524	1 NOTC_XENLA	P21783 xenopus lae
22	277	5.7	5179	1 MUC2_HUMAN	Q02817 homo sapien
23	276.5	5.7	3396	1 PGCV_HUMAN	P13611 homo sapien
24	273	5.6	1251	1 YQ03_CAEEL	Q09550 caenorhabdi
25	272.5	5.6	1075	1 FLO5_YEAST	P38894 saccharomyc
26	272	5.6	3358	1 PGCV_MOUSE	Q62059 mus musculu
27	269	5.5	1041	1 EGT2_YEAST	P42835 saccharomyc
28	267.5	5.5	2738	1 PGCV_RAT	Q9erb4 rattus norv
29	265.5	5.5	2109	1 PGCA_CHICK	P07898 gallus gall
30	265	5.5	995	1 YI09_YEAST	P40442 saccharomyc
31	264.5	5.4	881	1 YJH8_YEAST	P47033 saccharomyc
32	263.5	5.4	2700	1 ZAN_HUMAN	Q9y493 homo sapien
33	258.5	5.3	636	1 YNR6_YEAST	P53882 saccharomyc

34	257.5	5.3	860	1 CHI2_COCIM	P54197 coccidioide
35	255.5	5.3	676	1 MUC1_MESAU	Q00528 mesocricetu
36	255	5.2	605	1 YHC8_YEAST	P38739 saccharomyc
37	255	5.2	2437	1 NOTC_BRARE	P46530 brachydanio
38	252	5.2	2531	1 NOTC_MOUSE	Q01705 mus musculu
39	251	5.2	5376	1 ZAN_MOUSE	O88799 mus musculu
40	249.5	5.1	1150	1 APMD_PIG	P12021 sus scrofa
41	249	5.1	2531	1 NOTC1_RAT	Q07008 rattus norv
42	246.5	5.1	634	1 HWPI_CANAL	P46593 candida alb
43	246.5	5.1	2703	1 NOTC1_DROME	P07207 drosophila
44	244	5.0	862	1 PGCV_MACNE	Q28858 macaca neme
45	243	5.0	536	1 YEN1_SCHPO	O13695 schizosacch

ALIGNMENTS

RESULT	1
AMYH_YEAST	
ID	AMYH_YEAST
AC	P08640; P08068; STANDARD; PRT; 1367 AA.
DT	01-AUG-1988 (Rel. 08, Created)
DT	01-FEB-1995 (Rel. 31, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Glucosylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucohydrolase).
DE	STAL OR STA2 OR MAL5 OR YIR019C.
GN	Saccharomyces cerevisiae (Baker's yeast).
OS	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX	NCBI_TaxID=4932;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=S288C / AB972;
RA	Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA	Churche C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA	Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA	Louis E., Lyne G., Moule S., Moule T., Odell C., Pearson D.,
RA	Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA	Walsh S.V., Whitehead S.;
RL	Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RC	MEDLINE=87194600; PubMed=3106330;
RA	Yamashita I., Nakamura M., Fukui S.;
RT	"gene fusion is a possible mechanism underlying the evolution of STAL.";
RL	J. Bacteriol. 169:2142-2149(1987).
RN	[3]
RP	SEQUENCE OF 1-31 FROM N.A.
RC	STRAIN=SPX101-1C; PubMed=3141213;
RA	Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT	"Similar short elements in the 5' regions of the STA2 and SGA genes from Saccharomyces cerevisiae.";
RL	FEBS Lett. 239:179-184(1988).
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose.
CC	-1- SIMILARITY: TO S.POMBE SPCC215.13.
CC	-1- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; Z38061; CAA86176.1; -
DR	EMBL; M16164; AAA35014.1; -
DR	EMBL; M16165; AAA35015.1; -

DR EMBL; X13857; CAA32069.1; -.
DR PIR; B26877; B26877.
DR PIR; A26877; A26877.
DR PIR; S48478; S48478.
DR SGD; S0001458; MUCL.
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1367 GLUCOAMYLASE S1/S2.
FT DOMAIN 210 1367 SER/THR-RICH
FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;

Query Match 8.2%; Score 399; DB 1; Length 1367;
Best Local Similarity 24.3%; Pred. No. 2.5e-12;
Matches 196; Conservative 117; Mismatches 325; Indels 170; Gaps 30;

QY 4 TETVRSVAPMRGGEITAHLLTNSTT---SADVT-----GSSASYPEGYNASVLTQFSD 55
DB 400 STTESSAP-----VTSSTESSAPVTSSTTESSAPVTSSTTESSAPVTS 448

QY 56 STVQSGSHTALGDRYSSESS-----TSSSELSNAPRGERSIAGISYQVGRGTAEQR 111
DB 449 STTESSAPVTPSSSTTESSAPVTSSTTESSAPVTPSSSTTESSAPVTSSTTESS 508

QY 112 TSSDHTDHTYLSFTFKGERA-----LLSITDNSS---SSDIVESSTYIKISNSSHSEY 163
DB 509 SAPVTP-----SSSTESSAPAPTPSSSTTESSAPVTPSSSTTESSAPVTPSSSTTES 564

QY 164 SS--FSHAQTERNSISSYDGEYAQPTSPVLHTSNLPSTYPTIN-MPNTSVVLDTDAEF 220
DB 565 SSTRPVTSSSTESSA-----PVPTSSSTTESSAPVTPSSSTTESSAPAPTPSSS 617

QY 221 VSDSSSSSSSSSSSGGPPPLPVSQSHLFPSSLIPLSTRASVHLKSTSDASTPWSSS 280
DB 618 TTSSSAPVTSSTTESSAPVTPSSS-----TTSSSAPVTPSSSTTESSA 666

QY 281 PSLPVPVSLTT-SISAPLSVSQTLTPOSSSTPVLPRARET-----PVTSFOTSTMTSPMT 334
DB 667 PVTPPSSSTTESSAPVTSSTT---ESSAPVTSSTTESSAPVTPSSSTTESSAPV 723

QY 335 LHSQADLKSQSTPHQEKVITESKPSLSVLPTESTKA-----VTTNSPLP--- 381
DB 724 TPSSSTTESSAPVTPSSSTTESSAPVTSSTTESSAPVTPSSSTTESSAPVTPPS 783

QY 382 -----PSLTSTSTQTL-----PATSNLAQMSPTFTTILKTSQLMTTPGTL 425
DB 784 SSTTESSAPVTPSSSTTESSAPVTPSSSNITSSAPSPFPSSSTESSVPVTPPS 843

QY 426 SSTASLVGTGPIAVQTTAGKQLSLTHPEILVPOISTEGGISTERNRVIVDATTGLIPTSV 485
DB 844 SSTTESSAPVSSSTT-----ESSVAPVTPSSSN-----ITSSAPSIIPSSST 888

QY 486 PTSAKEMTKLGVTAEYSPASRLSGTSPSQTTW----- 520
DB 889 TEFSTGTCTTVPSSKYPGGQSTSVSTTETTVTPKTTTSVTPSTTTITTVTCSTGT 948

QY 521 ----STAEDLAPKSAFVAVOSSQSPPTLLSSASVNSCAVNPCLNGECVADNTSRGYHC 576
DB 949 NSAGETTSKSPKVTVTTPVTSTTSTSTTTTSTTTTTCVS---TGTSAGETTS--- 1001

QY 577 RCPSPS--WQGDGCDVDNECLNCPKPTATCNNTQGFICKCPV---GYOLE-KGICNLV 630
DB 1002 -CSPKTIITTVPGCSTSPSEFASE---STTSPPTPVTVVSTTVVTEYSTKPGGEIT 1057

QY 631 RTVFTEFKLRKLTINTVERHSDLOQEVEN--EITKTLNMFSAI-----PSYIR 677
DB 1058 TTEVTK-NIPTTVL-TTIAFTPSTVTTNFTPTTITTVTCSTGNSAGETTSKSPKVT 1115

QY 678 STVHASRESNAVISLQTTFSLASNVTL 705
DB 1115

DB 1116 TTVPCTGTGTGEYTTTATTLVTTAVTTTV 1143

RESULT 2
YMS6_YEAST
ID YMS6_YEAST STANDARD; PRT; 1140 AA.
AC Q04893;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 113.1 kDa protein in PRE5-FE14 intergenic region.
GN YMR317W OR YMR924.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z54141; CAA90835.1; -.
DR SGD; S0004936; YMR317W.
KW Hypothetical protein; Repeat.
SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

Query Match 8.1%; Score 396; DB 1; Length 1140;
Best Local Similarity 21.7%; Pred. No. 2.8e-12;
Matches 230; Conservative 161; Mismatches 387; Indels 284; Gaps 36;

QY 1 MSQTETVRSVAPMRGGEITAHLLTNSTTTSADVTG-----SSASYPEG-VNASVLQ 52
DB 52 VSSSTTLVSSVVP-----EFTSSSLSDTTASILSSELSVIFSLSYTSSDISSTVND 107

QY 53 FSDSTVQSGSHTALGDRYSSESSSTSSSELSNSA-----PRGERSIAGISYQVGRGTA 107
DB 108 VESTSGPSNSYALSSTNAQLSSSTTETDSISSAQTSSPQTSSNGGSSSEPLGKS 167

QY 108 IEQRTSDHTDHTYL-SSTFKGERALLSITDNSSSDIVESSTSYIKISNSSHSEYSSF 166
DB 168 SVLETTASSSDTTAVTSSTFT-----TLTDVSSSPKISSSGSAVTSVGTSDASKEVF 220

QY 167 SHAQTERNSISSYDGEYAQPTSPVLHTSNLPSTYPTINMPNTSVVLDTDAEFVSD--- 223
DB 221 SSSTSDVSSLLS-----STSSPA-----SSTSETLPFSTILSTISSVSEAP 265

QY 224 -----SSSSSSSSSSSSSGGPPPLPSPV-----SQSHLFSIL 257
DB 266 SATSSSVSSSEASSSTSSSVSEAPLATSSVVSSSEAPSTSSVSSSEAPSTSSSVSSEIS 325

QY 258 PSTRASV-----HLKSTSDASTPWS-----SSPPLPVSLTTSAP 295
DB 326 STTSSSVSSSEAPLATSSVVSSSEAPSTSSSVSSSEISSTSSSVSSSEAPLATSSVSSSEAP 385

QY 296 LSVSQTT---LQSSSTPV---LPRARETPVTSQTMTSTFMTLHSSQTADLKSGSTP 349
DB 386 SSTSSSVSSSEAPSTSSSVSEAPSTSSSVSSSEISSTSSSVSSSEAPSTSSSVSSSEAP 445

QY 350 HQEKVITESKPSLSVLPTESTKAVTTNSPLPPSLTSSSTEQTLPATSNLAQMSPTFTT 409
DB 446 SAISSLASSRLFSKKNSTVSTLVATEASSVTSSLRPSS--ETLASNIIESSLSTGYS 503

QY 410 TILKTSOPLMTPTG-----LSSTASLVGTPIAVOTTAGKQLSLT 449
 Db 504 TVSTTTSAASSTLGKSVSSNRWATSKTSTSSDLKSSVIFGNSVTTPSPASISLT 563
 QY 450 HPEILVPOISTEGIGSTERNRIVDATTGLIPL--TSVPTSAKEMTKL-----G 497
 Db 564 ASPL--PSVNSD-ITSSEASSISNLASSAPSDNNNTIASASLIVTKRNVSSIVSS 620
 QY 498 VTAEYSPASRLCTSPSPQTVTVSTADLAPKATFAVQSTQSPITLSSASVNSCAVN 557
 Db 621 ITSETTNESNLATSSYLSLUNKATASLSTSNAT-----SASNVPTTFSSMSHTSVIT 676
 QY 558 P-----CLHNGECVADNTRSGYHCRCPSPSWOGDGSDVDNECLSNPCPSTATCNCNTQ 610
 Db 677 PGFSTSSASLAINSTVSSSLAGYSFSTPES-----SPTTSLVTSFAPSTVSSMTTSA 730
 QY 611 SFICKCPGVQLEKIGKICNLVTRFTEPKLRTELNTVVEKHSDLOEVENEITKTLNMCFS 670
 Db 731 PFINN-----STSRAPSPSTASFITESSTSSIVPLASGDVT-----S 768
 QY 671 ALPSYIRSTVHASRESNAVISLQTTESLASNVTLFDLADR-----MOKCVNSCKSSAE 724
 Db 769 SLAANHUTTSAPSTSAQLVSKSTT---SSSILVTPRIDRSGNSSTPASRIATSLPNKTT 825
 QY 725 VCQLLGS-----QRRIFRAGSLCKRKSPCEDKDTISCTDLGVALCQCKGYFQFNKMDHS 780
 Db 826 FVSSLSTSAHARNIFNSTVLATAKQIETLTST----- 858
 QY 781 CRACEDGYLENETCMSPFLGGLNGCNP---YQLITVIVAAAGGGLLILGIALIVTC 837
 Db 859 -----VNCNPTPNYNTKTVISRE-----TTAIGTVTSC 889
 QY 838 ---CRKNKNDISKLP-----KSGD-----FQMSPYAEY 863
 Db 890 SGGCTNRKKTTLITIPDIDASTVTTTCEKEVTSSTSGDEAHTTSTKISNFETSFES 949
 QY 864 PKNPRSEMGREATEMHEN--GSTKNLLQMDTVYVSPTSVRNP 904
 Db 950 FKDMKTSQETKKAKPGSETVRSSSSFEKT-----SPTTKASP 987
 RESULT 3
 DAN4_YEAST
 ID DAN4_YEAST STANDARD, PRT; 1161 AA.
 AC P47179;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Cell wall protein DAN4 precursor.
 GN DAN4 OR YJRI51C OR J2223.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Scarcez T.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REGULATION.
 RX MEDLINE=21113168; PubMed=11160904;
 RA Cohen B.D., Sertil O., Abramova N.E., Davies K.J., Lowry C.V.;
 RT "Induction and repression of DAN1 and the family of anaerobic
 RT mannoprotein genes in Saccharomyces cerevisiae occurs through a
 RT complex array of regulatory sites.";
 RL Nucleic Acids Res. 29:799-808(2001).
 CC -!- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (potential).
 CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO THE SRP1 / TIPI FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 249651; CAA89684.1; -
 DR SGD: S0003912; DAN4.
 DR InterPro: IPR000992; SRP1_TIP1.
 DR Pfam: PFO0660; SRP1_TIP1; 1.
 DR PROSITE: PS00724; SRP1_TIP1; 1.
 KW Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 1146 CELL WALL PROTEIN DAN4.
 FT PROPEP 1147 1161 REMOVED IN MATURE FORM (POTENTIAL).
 FT LIPID 1146 1146 GPI-ANCHOR (POTENTIAL).
 SQ SEQUENCE 1161 AA; 118358 MW; 7954C15D69F0CA58 CRC64;
 Query Match 7.1%; Score 343; DB 1; Length 1161;
 Best Local Similarity 21.3%; Pred. No. 1.le-09;
 Matches 204; Conservative 147; Mismatches 392; Indels 216; Gaps 34;
 QY 3 QTEVTSRSVAP-----MRGGEIT-----AHWLLTNSTTSADVGTGSSASYPE 43
 Db 60 KTEIYPEIAAAVDYGDFTYRLTGISGDEVTRMITGVPTWYSTRLKPAI-----SSALSKD 115
 QY 44 GVNASVLTFQSDSTVQSGGSHLTALGD-RSYSESSSTSSSSSLNSAPRGERSIAGISYQ 102
 Db 116 GIYTAIPTSTTTTKSSTSTPTTTITSTTSTTSTPTTTSTTTSTTTSTTTSTTTST 175
 QY 103 VRGTAIQRTSSDHTDHT-----YLSSTFFKGERALLSITDNSSSDSIVESTSYKIKSN 157
 Db 176 TSTPTPTTSTTTSTTTSTTTSTTTSTTTSTTTSTTTSTTTSTTTSTTTSTTTSTTTST 235
 QY 158 SSHSEYSSFSHAQTERNSISSYDGEYAOQSPTE---SPVLHTSNLPSTPTIN-MPNTSV 212
 Db 236 TSTTQSTSTKSTPTTSTTTSTTTSTTTSTTTSTTTSTTTSTTTSTTTSTTTSTTTST 295
 QY 213 VLDTDAEFVSDSSSSSS-----SSSSSSSGPPLPLPSVQSQSHLFSILPSTRASVHLL 267
 Db 296 ---TSSTFTSSASASVISTATTATTFASLTTPATSTASTDHTTSSV---STTNAFTS 350
 QY 268 KTSDDASTPWSSSPPLVSLTSTTSTAPLSVSOFT--LPOSSSTPVLPRARETPTVSOT 325
 Db 351 ATTTTSDTYISSSP---SQVTSAAETTVSEVTSSVEPTRSSQVTSAAETTVSEFTS 407
 QY 326 STMTSEMTMLHSS---QTADLKSOSTPHQEKVITESKPSLVS-----LPTSTKAVT 375
 Db 408 SVEPTRSQVTSAAETTVSEFTSSVEPTRSSQVTSAAETTVSEFTSSVEFTSSQVTS 467
 QY 376 TNSLPPLPSTESSTEQTLPATSTNLQMSPTFTTILKTSQPLMTTPTGLS---STASLV 432
 Db 468 SAETTVSEFTSSVEPTRSSQVTSAAE--PTTVSEFTSSVEPTRSSQVTSAAETTVSEV 525
 QY 433 TGPVAVOTTAGKQLSLTHP-----EILVPQISTEGGISTERNRIVDA-- 475
 Db 526 TS---SVEPIRSSQVTTTEPVSSFGSTSEITSSAEPLSFASKATTSABESSNQITISSEL 583
 QY 476 -TTGLIPLTSVPTSAKEMTKLGVTAEYSPASRLCTSPSPQTVTVSTADLAPKASATFA 534
 Db 584 IVSSVITSSSEIPISSIEVLTFSSGSSSVEPTSL---VGPSSDESISSESISATSTFTA 640
 QY 535 VQSS-----TQSPPTLSSASVNSCAVN-----PCLHNGECVADNTRSGYHCR 578
 Db 641 VWSSKKADEFTRTVSAKSDVNSGNSSTQSTTFFATPTSTPLAVSTVYTSST----- 692
 QY 579 PPSMQGDCSDV--NECLSNPCPSTATCNCNTQSGTFCKCPGVQLEKICNLVTR--F 633
 Db 693 -----DSVSPNIPFSEISSPESSTA-ITSTSTSFIAE-----RTSSLY 730
 QY 634 VTEPKLRTELNTVVEKHSDLOEVENEITKTLNMCFSALPSYI--RSTVHASRESNAVI 691

QY 783 -ACEDGYRLE--NETMSCPFGGLGNCNPNYQLITVIVIAAAGGGLLILGIALIVTCR 839
Db 451 CVCMPGYQVRANGNRKCEPCPGYSGMCKDQFQLILVITGVIAGALLILLIAFVSARS 510
QY 840 KNKNDISKLFKSGDFQ 856
Db 511 KNKK-----KDGEQ 520

RESULT 5
PGCN_RAT
ID PGCN_RAT STANDARD; PRT: 1257 AA.
AC P55067;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurocan core protein precursor (245 kDa early postnatal core
DE glycoprotein) [Contains: 150 kDa adult core glycoprotein].
GN CSPG3 OR NCAN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RA MEDLINE=92406907; Pubmed=1326557;
RT "Cloning and primary structure of neurocan, a developmentally
RT regulated, aggregating chondroitin sulfate proteoglycan of brain.";
RL J. Biol. Chem. 267:19536-19547(1992).
RN [2]
RN CHARACTERIZATION.
RX MEDLINE=94230574; Pubmed=7513709;
RA Friedlander D.R., Milev P., Karchikeyan L., Margolis R.K.,
RA Margolis R.U., Grumet M.;
RT "The neuronal chondroitin sulfate proteoglycan neurocan binds to the
RT neural cell adhesion molecules Ng-CAM/L1/NILE and N-CAM, and inhibits
RT neuronal adhesion and neurite outgrowth.";
RL J. Cell Biol. 125:669-680(1994).
CC -1- FUNCTION: MAY MODULATE NEURONAL ADHESION AND NEURITE GROWTH DURING
CC DEVELOPMENT BY BINDING TO NEURAL CELL ADHESION MOLECULES (NG-CAM
CC AND N-CAM). CHONDROITIN SULFATE PROTEOGLYCAN; BINDS TO HYALURONIC
CC ACID.
CC -1- TISSUE SPECIFICITY: EARLY POSTNATAL AND ADULT BRAIN; NOT EXPRESSED
CC IN KIDNEY, LUNG, LIVER AND MUSCLE
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
CC O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC -1- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC
CC DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL
CC BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M97161; AAC37679.1; -
DR HSP; P00740; 1IXA.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF-2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR003599; Ig.

DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR000538; Link
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00008; EGF; 2.
DR Pfam: PF00047; Ig; 1.
DR Pfam: PF00059; lectin_c; 1.
DR Pfam: PF00084; sushi; 1.
DR Pfam: PF00193; Xlink; 2.
DR ProDom: PD000918; Link; 2.
DR SMART: SM00032; CCP; 1.
DR SMART: SM00034; CLECT; 1.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00001; EGF_Like; 1.
DR SMART: SM00409; IG; 1.
DR SMART: SM00445; LINK; 2.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00022; EGF_1; 3.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS01241; LINK; 2.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1257 NEUROCAN CORE PROTEIN.
FT CHAIN 639 1257 150 KDA ADULT CORE GLYCOPROTEIN.
FT DOMAIN 51 146 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 176 253 LINK 1.
FT DOMAIN 274 355 LINK 2.
FT DOMAIN 949 985 EGF-LIKE 1.
FT DOMAIN 987 1023 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1025 1154 C-TYPE LECTIN.
FT DOMAIN 1155 1213 SUSHI.
FT DISULFID 58 139 BY SIMILARITY.
FT DISULFID 181 252 BY SIMILARITY.
FT DISULFID 205 226 BY SIMILARITY.
FT DISULFID 279 354 BY SIMILARITY.
FT DISULFID 303 324 BY SIMILARITY.
FT DISULFID 953 964 BY SIMILARITY.
FT DISULFID 958 973 BY SIMILARITY.
FT DISULFID 975 984 BY SIMILARITY.
FT DISULFID 1029 1040 BY SIMILARITY.
FT DISULFID 1057 1149 BY SIMILARITY.
FT DISULFID 1125 1141 BY SIMILARITY.
FT DISULFID 1156 1199 BY SIMILARITY.
FT DISULFID 1185 1212 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 944 944 O-LINKED (XYL. . .) (CHONDROITIN SULFATE).
FT CARBOHYD 967 967 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1164 1164 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1257 AA; 135544 MW; 992B33DCFA19EE1B CRC64;

Query Match 6.88; Score 328; DB 1; Length 1257;

Best Local Similarity 22.08; Pred. No. 6.2e-09; Matches 155; Conservative 107; Mismatches 255; Indels 186; Gaps 31;

QY 1 MSQPTETSRVSVPARGGEITAHMLLTNTSTTSADVTGSSA-----SYPEGVNASVLTFQ 53
Db 418 LTRTQASQETLSTPPGGTTLASWLLTGTSTSGVPSLSGLVDMEETPSG-----TQV 471
QY 54 SDSTVQSGGSHALGDRSYSESSTSSSESLSNSAPRGERSIAGISYGVQVRGTAEIORTS 113
Db 472 APTPTMRGRFKGLNGRHFQQQG--PEDQLLEAAEASAPPTLEVTADHMGPSAATALE 529
QY 114 SDHTDHTYLSSTFTTKGERALLSITDNSSSDIVESSSTYKISNSSHSEYSSFSHAQTER 173
Db 530 SDOSHSPW-----AIL-----TNEVDVPGAGS-----LGRSLPESRKWSPSLISP 570

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Brain;
RX MEDLINE=96039250; PubMed=7490074;
RA Rauch U., Grimpe B., Kulbe G., Arnold-Ammer I., Beier D.,
RA Faessler R.;
RA "Structure and chromosomal localization of the mouse neurocan gene.";
RL Genomics 28:405-410(1995).
CC -1- FUNCTION: MAY MODULATE NEURONAL ADHESION AND NEURITE GROWTH DURING
CC DEVELOPMENT BY BINDING TO NEURAL CELL ADHESION MOLECULES (NC-CAM
CC AND N-CAM). CHONDROITIN SULFATE PROTEOGLYCAN; BINDS TO HYALURONIC
CC ACID.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAM/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X84727; CAA59216.1; .
DR HSP; P00740; IEDM.
DR MGD; MGI:104694; Cspg3.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR000538; Link.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00008; EGF; 2.
DR Pfam: PF00047; Ig; 1.
DR Pfam: PF00059; lectin_c; 1.
DR Pfam: PF00084; sushi; 1.
DR Pfam: PF00193; Xlink; 2.
DR ProDom: PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE-LECTIN_2; 1.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1268
FT DOMAIN 51 146
FT DOMAIN 176 253
FT DOMAIN 274 355
FT DOMAIN 960 996
FT DOMAIN 998 1034
FT DOMAIN 1036 1165

FT DOMAIN 1166 1224
FT DISULFID 58 139
FT BY SIMILARITY.
FT DISULFID 181 252
FT BY SIMILARITY.
FT DISULFID 205 226
FT BY SIMILARITY.
FT DISULFID 279 354
FT BY SIMILARITY.
FT DISULFID 303 324
FT BY SIMILARITY.
FT DISULFID 964 975
FT BY SIMILARITY.
FT DISULFID 969 984
FT BY SIMILARITY.
FT DISULFID 1040 1051
FT BY SIMILARITY.
FT DISULFID 1068 1160
FT BY SIMILARITY.
FT DISULFID 1136 1152
FT BY SIMILARITY.
FT DISULFID 1167 1210
FT BY SIMILARITY.
FT CARBOHYD 121 121
FT CARBOHYD 339 339
FT CARBOHYD 742 742
FT CARBOHYD 978 978
FT CARBOHYD 1175 1175
SQ SEQUENCE 1268 AA; 137200 MW; 3014E8E202A2FAEC CRC64;

Query Match 6.5%; Score 315.5; DB 1; Length 1268;
Best Local Similarity 23.0%; Pred. No. 2.5e-08;
Matches 164; Conservative 95; Mismatches 259; Indels 195; Gaps 31;

Qy 1 MSQETVSRVAPMRGGEITA-----HWLTNTSTTSADVTGSS-----ASYPEGVNA 47
Db 418 LTWTQAPETLPGTGGTFLASWPSSEKWLFTGAPSGMGVSPSDMGVDMETPLG--- 474
Qy 48 SVLQFSDSTVQSGSHPTALDRSYSESSSTSSSSSLNASSAPRGERSIAGSYGVGRGTA 107
Db 475 ---TQVAPTPTMRGRFRKGLNGRHFOQOGPEDQLDEV--AEPQAOPPLGATANHMRSA 529
Qy 108 IEQRTSSDHTYLSSTFTTKGERALLSITDNSSSDSDIVESSTSYKISNSHSHSYSFS 167
Db 530 ATEASESDQSHSPWAILTNEVDEPCAGSLGSRSLPESLMWSPS---LISPSVPS----- 580
Qy 168 HAQTERSNISSYDGEYAQPSTESPVLHTSNLPSYPTTNMNTSVVLTDAEFVSDSSSS 227
Db 581 ---TE-STPSPKPGAAEAPSVKSAIPLRLPSEPPEA---PSPG-----PSEALSA 624
Qy 228 SSSSSSSSSGPPPLPSPVSQSHLFF---SSILPSTRASVHLKSTSDASTPWSSSP-S 282
Db 625 VSLQASSADGSPDPPIVAMLRAPKWLPLRSTLVN-----LHSSQTA----- 341
Qy 283 PL-----PVSL-----TTTSAPLSVSQTTLPQSSSTPV---LPRARET 318
Db 671 PLPSWVPEEQAVRPVSLGAEDLETPFTTIAAPVEASHRS-PDADSIEIGTSSMRATKH 729
Qy 319 PVTSEFTSTWTSFWM-----LHSSQTA----- 341
Db 730 PISGFWASLDSNVTMPVPSPDAGILGTESGVLDPGSPGSGQATVEKVLATWLPQ 789
Qy 342 --DLKSQSTPHOEKVTESKPSLVSL-----PTESTKAVTNTSPLPPLSTESTEQT 392
Db 790 GLDQSGQSTPMEANGVAVSMEPT-VALEGGATEGMEATREV-----PSTADATWE--- 840
Qy 393 LPATSTNLQAQSPPTFTTILKTSQPLMTTPGTLSTSLASTLVTPGPIAVQTAGKQLSL-THP 451
Db 841 ---SESRSASSTHIAVTMARACGMPTLTSSBSGHPGPKGMVQAQESLEPLNTPSH 896
Qy 452 --EILVPOISTEGGISTERNRVIVDATGTL--IPITSVPTSAKEMTTKLGVTAEVSPASR 507
Db 897 WSSLVVPMDVEVASVSSGE-----PTGLWDIPSTLIPVSLGLDESVLNVAA----- 942
Qy 508 SLGTSPPQTTVSTAEADLPKSAFVQSSTQSTPTTLLSSASVNSCAVNCPLHNGECVA 567
Db 943 ----SPS-----VEGFWEV-----SGQEDPT-----DPCENNPLHGTC 977
Qy 568 DNTSRGYHCRPPSPWQGDGDCSDVNECLSNPCPSTATCNNTQGSFICKPCPVGY 620
Db 978 NGTV--YCCSCDQYAGENCEIDDDCLCSPCENGCTCIDVNGFICLCIFSY 1028

RESULT 11
AGAL_YEAST STANDARD; PRT; 725 AA.
ID AGAL_YEAST STANDARD; PRT; 725 AA.
AC P32323;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE A-agglutinin attachment subunit precursor.
GN AGAL OR YNR044W OR N3431.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=91304412; PubMed=2072914;
RA Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.;
RT "The AGAL product is involved in cell surface attachment of the
RT Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin.";
RL Mol. Cell. Biol. 11:4196-4206(1991).
RN [2]
RN SEQUENCE FROM N.A.
RA Pohl T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
CC SUBUNIT. S-CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
CC CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
CC RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
CC AGGREGATION DURING MATING.
CC -!- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
CC A CORE SUBUNIT.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M60590; AAA34382.1; -;
DR EMBL; Z71659; CAA96325.1; -;
DR PIR; S17031; S17031.
DR PIR; A41258; A41258.
DR SGD; S0005327; AGAL.
KW Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;
KW Pheromone response.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 725 A-AGGLUTININ ATTACHMENT SUBUNIT.
FT DOMAIN 53 493 2 X APPROXIMATE REPEATS.
FT REPEAT 53 149 1-1.
FT REPEAT 395 493 1-2.
FT DOMAIN 182 307 18 X APPROXIMATE TANDEM REPEATS, SER/THR-
FT RICH.
FT REPEAT 182 188 2-1.
FT REPEAT 189 195 2-2.
FT REPEAT 196 202 2-3.
FT REPEAT 203 209 2-4.
FT REPEAT 210 216 2-5.
FT REPEAT 217 223 2-6.
FT REPEAT 224 230 2-7.
FT REPEAT 231 237 2-8.
FT REPEAT 238 244 2-9.
FT REPEAT 245 251 2-10.
FT REPEAT 252 258 2-11.
FT REPEAT 259 265 2-12.
FT REPEAT 266 272 2-13.
FT REPEAT 273 279 2-14.

FT REPEAT 280 286 2-15.
FT REPEAT 287 293 2-16.
FT REPEAT 294 300 2-17.
FT REPEAT 301 307 2-18.
SQ SEQUENCE 725 AA; 73353 MW; 70420C853B0B01F8 CRC64;

Query Match 6.3%; Score 305; DB 1; Length 725;
Best Local Similarity 24.4%; Pred. No. 4.3e-08;
Matches 186; Conservative 97; Mismatches 246; Indels 234; Gaps 33;

QY 4 TETVRSVAP-----MRGEITAH--WL-LTNSTTSADVTSAGSYPEGVNASVLTQFSDS 56
DB 45 TTIVSPALVSTSTIVQAGTTTLTYTWCPLTVSTSSAAEISPSISY-----ATTLSREFTL 99
QY 57 TVQSGGSHALTGDRSYSESSTSSSLNSA--PRGERSIAGISYGOVRGTAIEQRTSS 114
DB 100 TL-----STEVCSEACPSSTLTPTTSLSVTSKFTSYICPTC----- 136
QY 115 DHTDHTVLSSTFTKGERALLSITD--NSSSDIDVSSSYIKISNSSHSEYSSFSHAQTER 173
DB 137 -HT--TAISLSEVGTTVTVSSSAIEPSSAIIIPVTTLSSTSSNPTTSLSTSTSP 193
QY 174 SNISSYDGEYAQPSTESPVLHTSNLPSTPTINPNSTSVVLDTDAEFVSDSSSSSSSSSS 233
DB 194 SS-----TSTSPST-----STSSSTSTSSSTSTSSS 222
QY 234 SSSGGPPLPSPVQSHHFLSSILPSTRASVHLKLTSDASTPWSSPSPLPVSLTSTTS 293
DB 223 STSTSP-----SSYSTSSSLTSSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 269
QY 294 APLSVSQTTLTPOSSSTPVLPRARETPTVTSQTSTMTSFTMTLHSSQTADLKSQSTPHQEK 353
DB 270 ---SSSTSTSPSSKST-----SASSTSTSYSTST----- 296
QY 354 VITESKSPSLVSLPTESTKAVTNSPLPPLSTESSTEQ-----LPATSTNLQMSPTF 407
DB 297 -----SPSL-----TSSSPILASTSPSTSTSTSTSTSTSTSTSTSTSTSTSTST 344
QY 408 TTTILKTSQPLMTTGGPLSTASLVGTPIAVQTGAKQLSLTHPEILVPOISTEGGISTE 467
DB 345 -----STPVISVPSTSSNNVATPMTSTVETTVSSQSS-----SEYITKSSIST- 388
QY 468 RNRVIVDATTGLPLTSPVPSAKEMTKLGVTAEYS---PASRSLGTS--PSPQTVV--- 519
DB 389 -----TIPFSMSMT---YFTTVSGVTMTTWTWCYSSSESTSTLTSMHETVTTD 434
QY 520 --VSTAEDLAPKSAFAVQSSQ-----SPTTLSSASVNSCAVNPCLHNGECVA 567
DB 435 ATVCTHESCHPSTQSTSLITSSIKMSTKNVATSVSTSTVESYACSTCAETSHSYVQTA 494
QY 568 DNTSRGYHCRPPSWOGDDCSVDNECLSNPCPSTATCNNTQGSFICKPCPVGYOLEKGIC 627
DB 495 SSSSVTQQTSTKSW-----VSSMTTSDDEFNKHATGKYHVTSSGT 535
QY 628 NLVRTFVTEFKLRTFLNTTVEKHSLOEVENETTKTNCFSAIPSYIRSTVHASRESN 687
DB 536 STSTSTSVSE-----ATSTSSIDSESQEQSSHLST-----SVLSS---SSLSATLSSD 580
QY 688 AVVISLQTTESLA---SNVTFLDLADRMQKCVNSCKSSAEVCQ 727
DB 581 STILLSSVSSLSVEQSPVTLQIS-----STSEILQ 612

RESULT 12
MUC4_HUMAN
ID MUC4_HUMAN STANDARD; PRT; 610 AA.
AC Q99102;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mucin 4 (Tracheobronchial mucin) (Fragment).
GN MUC4.

OC Eukaryotaxis elegans.
OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RC WILKINSON-SPROAT J.;
RA WILKINSON-SPROAT J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.

```

CC |nls swiss-prot entry is copyrighted. it is produced through a collaboration
CC |between the Swiss Institute of Bioinformatics and the EMBL outstation
CC |at the European Bioinformatics Institute. There are no restrictions on its
CC |use by non-profit institutions as long as its content is in no way
CC |modified and this statement is not removed. Usage by and for commercial
CC |entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC |or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48544; CAA88444.1; -.
DR EMBL; Z48582; CAA88444.1; JOINED.
DR WormPep; ZK945.10; CE01732.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 11 30 POTENTIAL.
FT DOMAIN 273 546 SER/THR-RICH.
FT DOMAIN 656 752 SER/THR-RICH.
SQ SEQUENCE 796 AA; 84306 MW; 76DC5B03E6357A6A CRC64;

Query Match 6.0%; Score 293.5; DB 1; Length 796;
Best Local Similarity 23.8%; Pred. No. 1.7e-07;
Matches 182; Conservative 92; Mismatches 245; Indels 247; Gaps 36;

Qy 3 QTEVTSRVAPMRGGEITAHLLNTSTTSADVTGSSAS-----YPGCVNASVLTFQSDS 56
| | | | | : : : : : | : : : : | : : : : |
Db 130 QTET-----RLVGLFLNSTWITLNEVNDDEISIAVEAKYECYDDGIDR-----CDG 177

Qy 57 T---VQSGSGHTA-LGDRSYSESSSTSS--SESLSNAPPARGERSIA-----GISY-QQ-- 102
| : | | : | : | : | : | : | : | : | : | : |
Db 178 SLWLQVGGNEWALLGYREKCESGEINEEYARRMKRPYRSEKSTALSDSQGVYDQVQL 237

Qy 103 ---VRGTALEQRTSDSDTHDHYLLSTFTFKGERALLSIDTNSSSSDIVESSSYIKINSNH 160
| | | | | | | | | : : : : : | : | | : |
Db 238 KGVRAKQFSMTSG-----SPTLRMRK---DAGDNTCDYTIESTSTSS----- 277

Qy 161 SEYSFSHAQTERSNISSYDCGEYAQPST---ESPVLHTSNLPSTYPTPI-----NMPN 209
| : | : | : | : | : | : | : | : | : | : | : |
Db 278 -----TTPTPTTVTSIVTSTTVPTSTIVTAMSTSTSTPS 315

Qy 210 TSVVLD-TDAEFVSDSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 268
| | : : | | | : | : | : | : | : | : | : | : | : |

```


DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP).
GN CP5G2.
OS Bos Taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS VO; V1; V2 AND V3).
RC TISSUE-Forebrain;
RX MEDLINE=98288320; PubMed=96241174;
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain.";
RL J. Biol. Chem. 273:15758-15764(1998).
RN [2]
RP SEQUENCE OF 21-53; 78-96; 226-250; 262-277; 295-306; 314-324; 329-331
RP AND 342-348.
RC TISSUE-Spinal cord;
RX MEDLINE=92062692; PubMed=1720020;
RA Perides G., Biviano F., Bignami A.;
RT "Interaction of a brain extracellular matrix protein with hyaluronic
RT acid.";
RL Blochim. Biophys. Acta 1075:248-258(1991).
CC -1- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS: At least 4 isoforms; V0 (shown here), V1, V2
CC and V3; are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed
CC in the central nervous system, and in a number of mesenchymal and
CC epithelial tissues; the major isoform V2 is restricted to the
CC central nervous system.
CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (by similarity).
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF060456; AAC24358.1; -;
DR EMBL; AF060457; AAC24359.1; -;
DR EMBL; AF060458; AAC24360.1; -;
DR EMBL; AF060459; AAC24361.1; -;
DR HSSP; P01132; 1EGF.
DR InterPro: IPR00152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR011881; EGF_Ca.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003600; Ig_like.
DR InterPro: IPR000538; Link.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR InterPro: IPR001304; Lectin_c.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.

DR Pfam; PF00193; xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00409; IG; 1.
DR SMART; SM00410; IG_like; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 3381 VERSICAN CORE PROTEIN.
FT DOMAIN 37 138 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 168 245 LINK 1.
FT DOMAIN 266 347 LINK 2.
FT DOMAIN 349 1336 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
FT DOMAIN).
FT DOMAIN 1337 3074 GAG-BETA.
FT DOMAIN 3074 3110 EGF-LIKE 1.
FT DOMAIN 3112 3148 EGF-LIKE 2, CALCIUM-BINDING.
FT DOMAIN 3161 3275 C-TYPE LECTIN.
FT DOMAIN 3280 3338 SUSHI.
FT DISULFID 44 131 BY SIMILARITY.
FT DISULFID 173 244 BY SIMILARITY.
FT DISULFID 197 218 BY SIMILARITY.
FT DISULFID 271 346 BY SIMILARITY.
FT DISULFID 295 316 BY SIMILARITY.
FT DISULFID 3078 3089 BY SIMILARITY.
FT DISULFID 3083 3098 BY SIMILARITY.
FT DISULFID 3100 3109 BY SIMILARITY.
FT DISULFID 3116 3127 BY SIMILARITY.
FT DISULFID 3121 3136 BY SIMILARITY.
FT DISULFID 3138 3147 BY SIMILARITY.
FT DISULFID 3154 3165 BY SIMILARITY.
FT DISULFID 3182 3274 BY SIMILARITY.
FT DISULFID 3250 3266 BY SIMILARITY.
FT DISULFID 3281 3324 BY SIMILARITY.
FT DISULFID 3310 3337 BY SIMILARITY.
FT CARBOHYD 57 57 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 817 817 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1017 1017 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1333 1333 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1393 1393 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1437 1437 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1463 1463 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1653 1653 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1974 1974 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2045 2045 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2074 2074 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2103 2103 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2263 2263 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2290 2290 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2356 2356 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2623 2623 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2641 2641 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2919 2919 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 3052 3052 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 3354 3354 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 3364 3364 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARSPLIC 349 349 P -> R (IN ISOFORM V1 AND V3).

FT	VARSP LIC	350	1336	MISSING (IN ISOFORM V1).
FT	VARSP LIC	1337	3074	MISSING (IN ISOFORM V1).
FT	VARSP LIC	350	3074	MISSING (IN ISOFORM V3).
FT	CONFLICT	25	25	MISSING (IN REF. 2).
FT	CONFLICT	51	51	MISSING (IN REF. 2).
FT	CONFLICT	89	89	N -> D (IN REF. 2).
FT	CONFLICT	96	96	Q -> D (IN REF. 2).
FT	CONFLICT	346	346	C -> R (IN REF. 2).
SO	SEQUENCE	3381	AA; 369984 MW; F09716FA7778D459	CRC64;
Query Match 6.0%; Score 293; DB 1; Length 3381;				
Best Local Similarity 22.0%; Pred. No. 9.4e-07;				
Matches 149; Conservative 108; Mismatches 274; Indels 146; Gaps 27;				
Qy	1	MSQETVSRSVAPMRGGEITAHWLLTNSTTSDVTGSSASYPE--GVNASVLTOFSDSTV	58	
Db	2555	LTMTSTILEIP---ELT-----SKNTVIDIDHTKPIYEDILGMOTDLDPEVPSP	2605	
Qy	59	QSGSHTALGDRSYSESSSTSSSESLNAPPGERSTAGISYGQVRGTAIEORTSSDHTD	118	
Db	2606	DSEESTQVQEKYCAANVLSSTENFEAS---GDILLANYQATPESKAPEDRNPLDHTD	2662	
Qy	119	HTYLSSTFTKGERALLSTIDNSSSDIVESSTYIKI-SNSSHSEYSSFSHAQT-----E	172	
Db	2663	F-----IFTTG-----IPILSSETELDVLLPTATSLPIPSKSATVNPESKTEAKTLEDIFE	2713	
Qy	173	RSNISSVDGEYAQSTSPVLHTNLSYPTTINMPNTSVVLDTAEPVSDSSSSSSSSSS	232	
Db	2714	SSTLS--DGOAI--ADQSEVISTILGYLERTQN-----EDEAKKYVSPSPQPEFSSG	2760	
Qy	233	SSSSGGPPLPLPSVSOHHLFSSILPSTRASVHLKSTSDASTPKSSSPSPPLPVSLTST	292	
Db	2761	AEALTDPTPVVSGTYI-----LTAQSLTEADPMVEGARLPDSIDTST	2804	
Qy	293	SA-----PLVSQSTTLPOSSSTPVLPRARETPVTFTQSTMTSEMTMLHSQ	339	
Db	2805	VSAPSELLSQTPPPLSIHLGSGDSEHSDLPQSA--LPSTDASTPPVS-----SGE	2855	
Qy	340	TADLKSOSTPHQKVITESKPSLVSLPTSTKAVTINSPLPPLSTESTEQTLPATSTN	399	
Db	2856	LANIEATFKPSEEDFYITEPP---SLPPDPESEDESKPKLLEPTEASATELIAQEEIE	2912	
Qy	400	LAQMSPTFTTTILKTSQ-----PLMTTP-----GTL---SSTASLVTGP	435	
Db	2913	IFQNSDN-TTSVQVSGEAVKVPFPIETPEAEAIVTAASETKLEGATLRPHSTASVINGV	2971	
Qy	436	IAYOTTACKQLSLTHPEILVP-QISTEGGISTERNRVIVDATTGLIPLTSVPT-----	487	
Db	2972	EAGVVPQSPQTSERPTILSPLEISPE-----TQAALINGEDSTVAAPKQVPTRMDSNK	3027	
Qy	488	---SAKEMTTKLGVTAEYSPASRLGTSPPSPQTVVYSTAEDLAPKSATFAVQSTQSPT	543	
Db	3028	QATLSTTELNTELA-----TPSFPLLETS-NETSFLIGINEESVEGTAVYL-----	3072	
Qy	544	TLSSASVNSCAVNPCLHNGECVADNTSRGYHCRCPSPSQWGDGCDVDVNECLSNPCPSTA	603	
Db	3073	-----PGPDRCKMNPCLNGGTCYETEIS--YVCVCVGYSGDRCELDFDECHSNPCRNA	3125	
Qy	604	TCNNTQGSFICKCPVGY	620	
Db	3126	TCIDGNTFRCLCLPSY	3142	

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 14:28:36 ; Search time 41.29 Seconds

(without alignments)
3963.507 Million cell updates/sec

Title: US-09-840-746-1

Perfect score: 4859

Sequence: 1 MSQETVSRKSVAPMRGGEIT.....FPGQYNPFSIDSESRDDVF 946

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phase:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_rvirus:.*
16: sp_bacteriap:.*
17: sp_archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3409	70.2	654	Q9ULI3	Q9ul13 homo sapien
2	1315	27.1	245	Q9BSV0	Q9bsv0 homo sapien
3	392.5	8.1	3507	Q23587	Q23587 caenorhabdi
4	387.5	8.0	2232	P91365	P91365 caenorhabdi
5	364	7.5	2271	Q990Y4	Q990y4 staphylococ
6	363.5	7.5	957	Q14651	Q14651 homo sapien
7	363.5	7.5	1217	Q9URW9	Q9urw9 homo sapien
8	352.5	7.3	1236	Q9C105	Q9c105 schizosacch
9	348	7.2	1079	Q9N4S7	Q9n4s7 caenorhabdi
10	347	7.1	528	Q29071	Q29071 sus scrofa
11	343.5	7.1	547	P97881	P97881 rattus norv
12	339	7.0	786	Q21027	Q21027 caenorhabdi
13	337.5	6.9	901	Q9H195	Q9h195 homo sapien
14	336	6.9	3072	Q939N5	Q939n5 streptococc
15	334.5	6.9	2178	Q9KWR3	Q9kwr3 streptococc
16	334	6.9	3381	Q9KX33	Q9kx33 streptococc

17	331.5	6.8	3971	3	Q96WV6	Q96wv6 schizosacch
18	331	6.8	1203	5	Q9NSK0	Q9nsk0 caenorhabdi
19	326	6.7	957	4	Q9URN0	Q9urn0 homo sapien
20	325	6.7	849	3	P87107	P87107 saccharomyc
21	323	6.6	4776	16	Q97P71	Q97p71 streptococc
22	321	6.6	878	4	Q9GZ22	Q9gz22 homo sapien
23	319	6.6	878	4	Q9H307	Q9h3q7 homo sapien
24	317.5	6.5	877	4	Q9H306	Q9h3q6 homo sapien
25	317	6.5	1443	3	Q9Y8F2	Q9y8f2 candida alb
26	317	6.5	1795	5	O76894	O76894 drosophila
27	316.5	6.5	1325	5	Q9BKV7	Q9bkv7 leishmania
28	314	6.5	816	11	O70474	O70474 rattus norv
29	313.5	6.5	534	3	Q94317	Q94317 schizosacch
30	313	6.4	1802	3	Q04051	Q04051 saccharomyc
31	310.5	6.4	948	3	O74346	O74346 schizosacch
32	310.5	6.4	3570	4	Q99552	Q99552 homo sapien
33	310	6.4	961	3	Q92223	Q92223 emericeila
34	307.5	6.3	2586	5	Q9VTK8	Q9vtk8 drosophila
35	303.5	6.2	1805	11	O63661	O63661 rattus norv
36	301	6.2	3178	5	Q969D4	Q969d4 caenorhabdi
37	300.5	6.2	622	4	Q14881	Q14881 homo sapien
38	300.5	6.2	1275	5	O76602	O76602 caenorhabdi
39	297	6.1	1459	5	O17084	O17084 caenorhabdi
40	289	5.9	1001	3	Q05164	Q05164 saccharomyc
41	287	5.9	716	4	Q9NVE4	Q9nye4 homo sapien
42	286.5	5.9	1321	4	O14594	O14594 homo sapien
43	284	5.8	512	4	Q9H3R2	Q9h3r2 homo sapien
44	282	5.8	513	4	O43418	O43418 homo sapien
45	281	5.8	866	12	O39781	O39781 equine herp

ALIGNMENTS

RESULT 1

Q9ULI3 ID Q9ULI3 PRELIMINARY; PRT; 654 AA.
AC Q9ULI3;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE KIAA1237 PROTEIN (FRAGMENT).
GN KIAA1237.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL: AB033063; BAA86551.1; -
DR InterPro: IPR000152; Asx_Hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00008; EGF; 3.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00001; EGF_like; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
DR PROSITE: PS01187; EGF_CA; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
FT NON_TER 1
SQ SEQUENCE 654 AA; 70602 MW; 8F51D7F639E8B0CE CRC64;

Query Match

70.2%; Score 3409; DB 4; Length 654;

Best Local Similarity 100.0%; Pred. No. 1.3e-199;		Matches 654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	293	SAPLSVSQTTLPQSSSTPVLPRARETPTVTSQTSMTSEMTLHSSQTADLKSTPHQE	352
DB	1	SAPLSVSQTTLPQSSSTPVLPRARETPTVTSQTSMTSEMTLHSSQTADLKSTPHQE	60
QY	353	KVITESKSPSLVSLPTESTKAVTNSPLPSLTESSTEQTLPATSTNLAQMSPTFTTIL	412
DB	61	KVITESKSPSLVSLPTESTKAVTNSPLPSLTESSTEQTLPATSTNLAQMSPTFTTIL	120
QY	413	KTSOPLMTTPTGLSTASLVTPGPIAVQTTAGKQLSLHPEILVPOISTEGGISTERNRVI	472
DB	121	KTSOPLMTTPTGLSTASLVTPGPIAVQTTAGKQLSLHPEILVPOISTEGGISTERNRVI	180
QY	473	VDATTGLIPTSVPTSAKEMTKLGVTAEYSPASRSISLGTSPSPQTTVVSTAEDLAPKSAT	532
DB	181	VDATTGLIPTSVPTSAKEMTKLGVTAEYSPASRSISLGTSPSPQTTVVSTAEDLAPKSAT	240
QY	533	FAVQSSSTQSPSTLSSASVNSCAVNPCLHNGECVADNTSRGYHCRCPSPSQGGDCSDVDN	592
DB	241	FAVQSSSTQSPSTLSSASVNSCAVNPCLHNGECVADNTSRGYHCRCPSPSQGGDCSDVDN	300
QY	593	ECLSNPCPSTATCMTNTOGSEFCCKCPVGQLEKGCICNLVRTFVTEFKLRTFLNTTVEKHS	652
DB	301	ECLSNPCPSTATCMTNTOGSEFCCKCPVGQLEKGCICNLVRTFVTEFKLRTFLNTTVEKHS	360
QY	653	DLQVEVEITKTLNMFCSALPSYIRSTVHASRESNAVVISLQTFSLASNVTLFDLADM	712
DB	361	DLQVEVEITKTLNMFCSALPSYIRSTVHASRESNAVVISLQTFSLASNVTLFDLADM	420
QY	713	QKVNCKSSAEVQCLGSGRRIFRAGSLCKRKSPCEDKDTISICTDLGVALCOCKSGYF	772
DB	421	QKVNCKSSAEVQCLGSGRRIFRAGSLCKRKSPCEDKDTISICTDLGVALCOCKSGYF	480
QY	773	QFNKMDHSCRACEDGYRLENETCMSCPFGLGGLNCGNPYQLITVVIAGGGLLLIIGIA	832
DB	481	QFNKMDHSCRACEDGYRLENETCMSCPFGLGGLNCGNPYQLITVVIAGGGLLLIIGIA	540
QY	833	LIVTCCRKNKNDISKLFKSGDFQMSPYAEYKPNRQSGREAIEMHENGSTKNLLQMT	892
DB	541	LIVTCCRKNKNDISKLFKSGDFQMSPYAEYKPNRQSGREAIEMHENGSTKNLLQMT	600
QY	893	DVYYSPTSVRNPELERNGLYPAYTGLPGSRHSCIFPGQYNPFSIDSSRRDYF	946
DB	601	DVYYSPTSVRNPELERNGLYPAYTGLPGSRHSCIFPGQYNPFSIDSSRRDYF	654
RESULT 2			
Q9BSVO	PRELIMINARY; PRT; 245 AA.		
ID	Q9BSVO		
AC	Q9BSVO;		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	HYPOTHETICAL 27.3 KDA PROTEIN (FRAGMENT).		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-OVARY, ADENOCARCINOMA;		
RA	Strausberg R.;		
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC004539; AAH04539.1; -.		
KW	Hypothetical protein.		
FT	NON_TER 1		
SQ	SEQUENCE 245 AA; 27334 MW; OCCAFBE55822EF47 CRC64;		

Query Match

Best Local Similarity 100.0%; Pred. No. 1.1e-72;

27.1%; Score 1315; DB 4; Length 245;

		Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	705	LFDLADRMQKCVNSCKSSAEVQCLGSGRRIFRAGSLCKRKSPCEDKDTISICTDLGVAL	764
Db	4	LFDLADRMQKCVNSCKSSAEVQCLGSGRRIFRAGSLCKRKSPCEDKDTISICTDLGVAL	63
Qy	765	COCKSGYGFQFNKMDHSCRACEDGYRLENETCMSCPFGLGGLNCGNPYQLITVVIAGGG	824
Db	64	COCKSGYGFQFNKMDHSCRACEDGYRLENETCMSCPFGLGGLNCGNPYQLITVVIAGGG	123
Qy	825	LLLILGIALIVTCCCRKNKNDISKLFKSGDFQMSPYAEYKPNRQSGREAIEMHENG	884
Db	124	LLLILGIALIVTCCCRKNKNDISKLFKSGDFQMSPYAEYKPNRQSGREAIEMHENG	183
Qy	885	TKNLQMTDVYYSPTSVRNPELERNGLYPAYTGLPGSRHSCIFPGQYNPFSIDSSRRD	944
Db	184	TKNLQMTDVYYSPTSVRNPELERNGLYPAYTGLPGSRHSCIFPGQYNPFSIDSSRRD	243
Qy	945	YF 946	
Db	244	YF 245	
RESULT 3			
Q23587	PRELIMINARY; PRT; 3507 AA.		
ID	Q23587;		
AC	Q23587;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	2K783.1 PROTEIN.		
GN	2K783.1.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Nematoda; Chromadorea; Rhabditida; Rhabditidae;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RX	MEDLINE=94150718; PubMed=7906398;		
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,		
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,		
RA	Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,		
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,		
RA	Jones M., Kersey J., Kirsten J., Laister N., Latreille P.,		
RA	Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,		
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,		
RA	Saldon N., Smith A., Sonhammer E., Staden R., Sulston J.,		
RA	Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,		
RA	Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;		
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.		
RT	elegans."		
RL	Nature 368:32-38(1994).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RA	Favello A., Vaudin M.;		
RT	"The sequence of C. elegans cosmid 2K783."		
RL	Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RA	Waterston R.;		
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; U13646; AAC24418.1; -.		
DR	HSSP; P00736; IAPQ.		
DR	InterPro; IPR000152; Asx_hydroxyl.		
DR	InterPro; IPR000561; EGF-like.		
DR	InterPro; IPR001881; EGF_Ca.		
DR	InterPro; IPR001254; Trypsin.		
DR	InterPro; IPR001507; zona_pellucida.		
DR	Pfam; PF00008; EGF; 15.		
DR	SMART; SM00179; EGF_CA; 12.		

D	b		2956	SLATCQSGTGVCIICRDGFTGDGTACSKKSTADCTSLPSLCADKAACDNSTRS--CECDA	3011
Q	y		581	SWOGDD--GSVDVNECL--SNPCPSTATCNNTQ-----	609
D	b		3014	GYIGDGYVCSPHPDCDLVRNLNCSPEAVCONRRCCQLPGFTGDGVKCYSIHERASNCSCQ	3073
Q	y		610	-----GSFICKCPVGYLEKGIC-----NLV-----	630
D	b		3074	DANAHCVGTTCKTCPGY-FGNGLCCVPDPPLCDVHVTGICHPNVAVNCNESRQCOCSSGFS	3132
Q	y		631	RTVFTEFKLRTELNTVEKHSDLOEVENIEIKTLNMCM-----FSALPSYSTSVSHASRE	685
D	b		3133	NGVSCFPQKSCTDSV-----CAKNAILPTGCICIRHGFGDFPYKCSSLVSKEP	3185
Q	y		686	SNNAVISLQTTFSLASNVTFLDLARMQKCVNSKSSRAEV-----LLQSQRIFRAG	739
D	b		3186	ANQQ-----DLSD-VSSCSVTPCDSLTQLCATISGEICKSGFRRNSTLSG	3227
Q	y		740	S-----LCKRKSPCEDKTICTLDLGVALCOCKSG-----YFOFNKMHDHSR	782
D	b		3228	SETCADIDECAEKSHKCDR-VATCRNFTGFHVCTCPDHGVGDITCVPHVNQGLKSVYCE	3286
Q	y		783	ACEDGYRL-----ENETCMSCPFGLGLGCCNPYLQILTVIAAAGG	823
D	b		3287	A--DGMTVLNGNETADFEGKIFVKGOAENPKCYKSFSSL--LNSRKPY-VEKVVFQHCDV	3341
Q	y		824	GLL-----LIIGIALIVTCRCNNKNDISKLIKFGSGDFQMSPSYAEYPKNPRSQEWGR	874
D	b		3342	QLLDNHMTASTVVQVKHAMFLT----NKAD-----SYDLR---CQYP-----IGS	3379
Q	y		875	EATEMHENGSGS	884
D	b		3380	RAVESHWNVSGS	3389
R	E	s	RESULT	4	
P	p	i	ID	P91365	PRELIMINARY; PRT; 2232 AA.
A	C		P91365;		
D	T		01-MAY-1997	(TREMBLrel. 03, Created)	
D	T		01-OCT-1999	(TREMBLrel. 10, Last sequence update)	
O	T		01-OCT-2000	(TREMBLrel. 15, Last annotation update)	
D	E		K06A9_1	PROTEIN.	
G	N		K06A9_1.		
O	S		Caenorhabditis elegans.		
O	S		Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;		
O	X		Rhabditiadae; Pelodierinae; Caenorhabditis.		
O	X		NCBI_TaxID=6239;		
R	P		[1]		
R	P		SEQUENCE FROM N.A.		
R	C		STRAIN=BRISTOL N2;		
R	L		Geisel C., Gattung S.;		
R	L		Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.		
C	C	-	ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY		
C	C	-	ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT		
C	C		OF THE A FORM.		
D	R	EMBL;	U80846;	AAC70889.1;	
D	R	EMBL;	U80846;	AAC70890.1;	
K	W		Alternative splicing.		
F	T	VARSPLIC	842	866	
F	T	VARSPLIC	870	1051	
F	T				VASSPPAPSTSQNPNPNTSSGSMI -> LATTSAKPSPVTVM CLFMVD (IN ISOFORM B). PYPSTSPVESSTTPSPGPSQPTTLTSTSPSPSOSTTIGST OGSTSPGISSTTSSEWTSGSQTPGSGSVTGPSTVSDST SGSTVTVGSTEGBSSSPISTSONTNFSTSGSGSMSTQTFQ SSQSTSPVESTSGATSSGSPGTTLSISPSPSPSTIGS SGSTSPVVISITSGSTE -> KEIQDTAINTKYTFNFAL LVASKLNESILTGYIONFYGESAGLNDHQYPTDYNGIKS VPPFDIGTDIDDLDKDVKSLATADTWPVADOTCMIFI SNAPEDEGGTTIKSTYTVEVTGVGLVGAKSIPIGLSIDK NIIVETGMNDORDASVVSKLEULELPTA (IN ISOFORM B).
S	Q	SEQUENCE	2232 AA;	213840 MW;	08D69FA63BE14CC8 CRC64.

Query Match 8.0%; Score 387.5; DB 5; Length 2232;
Best Local Similarity 22.8%; Pred. No. 5.2e-15;
Matches 233; Conservative 145; Mismatches 395; Indels 251; Gaps 34;

QY 4 TETVSRVAPM---RGGEITAHMLLTNSTGSAVDTGSGASYPEGVNASVLTFQSDSTVQS 60
DB 1276 TPNPQSSTPVVSTTTGEMTSH-----GSTQTPSTGTVTQPSNVSGS---NSSGSTVTI 1328
QY 61 GGSHTALGDRSYSES-----SSTSS-----SESL 84
DB 1329 GSSSASTSGSFKTSPSSISPVPTSSPIPTTRASTSGSTISDVSVSTTSIAPLSSSL 1388
QY 85 NSSAPGERIAGISYQVGRGA---IQRTSSDHTD-----HTYLSSTFT-KGERALL 134
DB 1389 PSTVPSSTQSFSTFEGSSKASSPVPSQTSTPTNTGTSTESSTLSSSTISGSTQHTM 1448
QY 135 SITDNSSSDIVESST--SYKISNNSHSEYSSSHAQTE---RNNISSYDGEVAQPTES 190
DB 1449 SKASSGSTSPSTNSQTGSTVTMGSSSTSGVSTSSASTQPMSTSQGSSAGSTVASSTAS 1508
QY 191 PVLHTSNLPSVTPTINPNPNTSVVLDTDAEFVSDSS-----SSSSSSSSSSSSG 238
DB 1509 PAA--SSTAPSTGMTSSTSGTGSTISESTTASASSQTGSTVTMGSSSTSGVSTSSAS 1567
QY 239 PPLPLPSVQSHHLLFSSILPSTRASVHLLKSTSDASTPWSSP----SPLPVSLLTSTSA 294
DB 1568 STQPMSTSQGSSAGSTVASTAGLVSTVPSSTGMTSGTSSCTVSGTISESTTASAS 1627
QY 295 PLSVSTQLPQSSSTPVLPRARETPTVTSFQTSMTSFTWMTLHSSQTADLKSQSTPHQKV 354
DB 1628 SQTGSTVTMGSSSTSGVSTSSASTQPMSTSQGSSAGSTVASTTGLVSTVPSSTGT 1687
QY 355 ITESKPSLSVLPSTKAVTNSPLPPLSTESTETLPTATST-----NLAQMSPT 406
DB 1688 MGTSSSTGTGSTISESTTASASASTQGSTVTMGSSSTSGVSTSSAGSQPMSTSQGSSA 1747
QY 407 FTTTILKTSOPL--MTTP---GTLSSSTASLVTC-----PIAVQTTA--GKOLSLTHPEI 453
DB 1748 GSTVVSSTASPAASSTAPSTGTMSSSTSGTGTMSQSSTAASTSTHTGTVTLGSSST 1807
QY 454 LVPOISTEGGISTERNRVIIVDATTGLPLTSVPTSAKEM--TTKLK-----VTAEY 502
DB 1808 SSNMSTSQGSSV--GSTVASSTAGLVSTVTPSPSTGMTSGTSGTISESTTASA 1865
QY 503 SPASRSLSGTSPTQTTVVST-----AEDLAPKATFAVQGST- 539
DB 1866 SSQTGSTVTMGSSSTSGVSTSSASTQPMSTSQGSSAGSTVASTAGLVSTVTPSPSTG 1925
QY 540 ----QSPTTLSSASVNSAVNPNLHNGECVADNTSRGYHCRCP-----PSWQGGD 586
DB 1926 TMGSTSGTGTGSTISESTTAASTSSQTGSTVTIGSTGTPSPSPRSLSQITITPSPSQST 1985
QY 587 CSVDVNECLNPNCPSTATCNNTQGSFCKCPGVQLEKGCINLVRTVTEFKLRFTLNT 646
DB 1986 ESTQTSPLSPSSPSPSTSHVSSEGT-----TMSSGATSGDKMPSLSS 2028
QY 647 TVEKHSDLQEVENEITKTLNCFESALPSYIRSTVHASRESNAVVISLQTTFSLASNVTLF 706
DB 2029 T-----GTVVSFSSRGS-----SLATTSAKPSVTCIL 2055
QY 707 DLADRMQK-----CVNSCKSSAEVCQLLGSORRI-----FRAGSLCKRKSPES 749
DB 2056 FMVDTQSKIEDQTAINTKYTFENFALLVASKLNNESTLGTIDNFGYSAGLNDHQVYPTD 2115
QY 750 D-----KDTISICTDLGVALCOCKSGYFOFNKMDHSCRACEDGYRLENETCM 796
DB 2116 DYNGIKSVPPFDGTDIDDLKDV-----DRSLATADWTPPVADQTCM 2160
QY 797 ----SCPEFG--LGLNLCGNPNQLITVWTAAGGGLLLLTGTA----LVITCCRKNKNDISK 847
DB 2161 IFISAPEDEYGGTTIKSTTYFETVGVVLVGGAKSIPGLSIDKNIVITNTMNDRDASA 2220

QY 848 LIFK 851
DB 2221 VWSK 2224

RESULT 5
Q99QY4 PRELIMINARY; PRT; 2271 AA.

AC Q99QY4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE SA2447 PROTEIN (HYPOTHETICAL PROTEIN SAV2654).
GN SA2447 OR SAV2654.
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879, 158878;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S. aureus (strain N315), and S. aureus (strain Mu50);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekinizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003138; BAB43752.1; -;
DR EMBL: AP003366; BAB58816.1; -;
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 2271 AA; 227844 MW; 7C2A7040D6C8289D CRC64;

Query Match 7.5%; Score 364; DB 16; Length 2271;
Best Local Similarity 22.6%; Pred. No. 1.4e-13;
Matches 183; Conservative 151; Mismatches 363; Indels 114; Gaps 22;

QY 2 SQTETVSRVAPMRGGEITAHMLLTNSTTADV-----TGSSASYPEGVNASVLTFQS 54
DB 1054 SDSKSLSTSQSGSTS-----TSTTSASVRTSSQSTSGSMSASQSDMSISSTFS 1108
QY 55 DSTVQSGSHALTALDRSYSESSTSSSESLNS-----SAPGERIAGISYQVGRGAIEQ 110
DB 1109 DSTSDSKASTA--SSEISQSASTSTSGSVSTSTSLSTNSERTSTSVSDSTSLSTSED 1167
QY 111 RTSDDHTDHYLSTFTFKGERALLSTDNSSSDI--VESSTSYTKIS---NSSHSEYSSF 166
DB 1168 SISESTSDSISEAISASESTISLSSESNSTSDSESQSAFLSLSSESTSESTSESV 1227
QY 167 SHAQTERNSLTSSVDGEVAQPT-----ESPVLHTSNLPSYTPPTINMP--NTSVVLD 216
DB 1228 SSSTSESTSLSDSTSESGSTSTSLNSTSGSASISTSESTSTFKSESVSTSLSMST 1287
QY 217 DAEFVSDSSSSSSSSSSSSGGPPLPLPSVQSQHHLLFSSILPSPRASVHLLKSTSDASTP 276
DB 1288 STSLNSTSLSTSLSDSTSDSKSLSTSMSTSDSISTSDSISTSLSTSLSGSTSESED 1347
QY 277 WSSS-----PSPLPVSLTSTTAPLSVS--QTLLPQSSSTPVLPRAR-----ETPVTSF 323
DB 1348 STSSSEKSDSTSMISMSQSTSGSTSTSTSLSDSTSTSLSLASMNQSGVDSNSASQ 1407
QY 324 QTSMTSFTWMLHSSQ--TADLKSQSTPHQKVITES--KSPSLVSLTSTKAVTTNSPL 380
DB 1408 SASNSTSTSESDSQSTSTYTQSTSQSESTSTSTSLSDSTSIKSTSQSGSTSTASL 1467

Oy	381	-----PPSLTESTEQTLPAATSNLAQMSPFTTTTILKTSOPLMTPTGTLSSSTASLV 432
Db	1468	S GSESEDQSQISTSAESTSESASTLSLSDSTSGNSGSASTSTSLNSASASBESDSST 1527
Oy	433	TGPITAVOTTAGKQLSLTHPEILLVPOILTSEGIGISTERNRVIVDATGTGLIPLTVPTSAKEM 492
Db	1528	S--LUSDSTSASMOSSESDDSQSTASLSDSLSTST-S-NRMSTIASLTSVSTSESGSTSES 1584
Oy	493	TTKLGVTF----AYSPASRSLGTPSPQTVVYTAEDLPKSAFVAQOSSQPSTLLSSS 548
Db	1585	TSESDSTSTSLSDSQSTSRGSTSASSGASTSTSTDSTRSTASTSTSMRTSTDQSQSWLSL 1644
Oy	549	ASNVCAPNCLHNGECVDANTSRGYHCRCPPSWGDCDCCSYDVNECLS-NPCPSTATCNN 607
Db	1645	TS-TSTSMSDSTSLSDSVSDSTS-----DSTSASTSGSNSVSISLSDSTSTS 1690
Oy	608	TOGFICKCPVGQLBKGCINLVRFVTVEFKLKRTLNTTVEKHSHDLQEVENETTKTLNM 667
Db	1691	TSASEVMSAgi-----SDSQMSESVNDSESV 1717
Oy	668	CFSALPSYIRSTVHASRESNAVVISIQTTF-----SLASNVTLFDLADRMQKVNSCKASA 723
Db	1718	SENSESDSKMSGSTSVSDGSLSVSTSLRKSESVSSLGSGSQMSDSVSTSDSSL 1777
Oy	724	EVCOLLGSORRIFRAGSLCKRKSPECXDKDTs 754
Db	1778	SVSTSLASSESVSESDSLDSKSTSGSTSTS 1808
RESULT	6	
O14651		
ID	O14651	PRELIMINARY; PRT; 957 AA.
AC	O14651:	O14650;
DT	01-JAN-1998	(TREMBLrel. 05, Created)
DT	01-JUN-1998	(TREMBLrel. 08, Last sequence update)
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)
DE	MUCIN	(INTESTINAL MUCIN) (FRAGMENT).
GN	MUC3.	
OS	Homo sapiens	(Human).
OC	Eukaryota; Metazoa; Chordata;	Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria;	Primates; Catarrhini; Hominiidae; Homo.
RN	NCBII_TaxID=9606;	
OX	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=97476275;	PubMed=9334251;
RA	Gum J.R. Jr., Ho J.J.-L., Pratt W.S., Hicks J.W., Hill A.S.,	
RA	Vinall L.E., Robertson A.M., Swallow D.M., Kim Y.S.;	
RT	"MUC3 human intestinal mucin. Analysis of gene structure,	
RL	terminus, and a novel upstream repetitive region.";	
J Biol. Chem.	272:26678-26686(1997).	
DR	EMBL; AF007194; AAC02272.1; -;	
DR	EMBL; AF007196; AAB84383.1; -;	
DR	EMBL; AF007195; AAB84382.1; -;	
DR	InterPro; IPR000561; EGF-like.	
DR	SMART; SM00181; EGF; 1.	
DR	PROSITE; PS00022; EGF_1; UNKNOWN_2.	
DR	PROSITE; PS01186; EGF_2; 1.	
KW	EGF-like domain; Glycoprotein.	
FT	NON_TER	1
SEQ	SEQUENCE	957 AA; 99554 MW; E6772256D600A1F CRC64;

Query Match	7.5%	Score 363.5;	DB 4;	Length 957;
Best Local Similarity	24.2%	Pred. No. 5.2e-14;		
Matches 240;	Conservative 124;	Mismatches 333;	Indels 295;	Gaps 49;

QY	1	MSQETVTSRS---	VAPMRGGEITAHWL-	LTNSPTSDVTGSSASYPEGVNASVLTFQ-	53
		: : : : : :	: : : : : :	: : : : : :	
Db	52	ITTTTSHSTPSTSSRTTETVSYSTPSTSSNWTETT--	SHSTPSYITSTTTTP		109
		: : : : : :	: : : : : :	: : : : : :	
QY	54	SDSTVQSGGSHALGDRSYSE---	SSSTSSSESLSNSSAPRGSRISAGISYGOVGRGTAIEQ		110
		: : : : : :	: : : : : :	: : : : : :	

Db	110	SSSTPSPSSSIITTTTSHSTPGTSSITTTTSHSTPGTSSII-----154
Qy	111	RTSDHDHTYLSSTFTKGERALLSIIDNSSSSDIVESSTSYI----KISNSHSEYSFF166
Db	155	-TTTETTSHD--TSPFS-----SIITSETPSHSTPSSSTSLITTKTTSHTSPSTSSI205
Qy	167	SHAOTERNISSYDGEYAQSPSTESPVLIHTNLPSYTPPINPNT---SVVLDTDAEFVSD223
Db	206	TTTETTHSARSFTSSITTTETTS-----HNRSFTSSITTTETNSHSTTSFTSSITTE260
Qy	224	SSSSSSSSSSSSSGPPLP-----LPSVSQHHLFSSILPSPRASVHLKSTSDA----273
Db	261	TTSHSTPSSFSSITTTETPLHSTGLPSPWTTTKTTSIIITPGTSLITTTTETTHSTPGF320
Qy	274	-----STPMSSSPPLSVLTSTTSAPLSVSQTLTPOSSSTPVLPRARET--P319
Db	321	TSSITTTETTSSTP-SLSSTIYSTVSTTALTSHFTSETAVPTPTPSSLSTDIP379
Qy	320	VTSQTSTMTSFMTHLHSSQTADLKSOST-----PHOEKVITESKSPSLSLPTES---370
Db	380	TTSLRLTLPSPVGSTSLITTTTDPSPITDITSLPTRTHII--SSSPISQSTETSSLVGT437
Qy	371	-----TKAVTNSPL-----PPSLTSSSTEQTLPAINST-----398
Db	438	TSPTMTSVRMFLRITENTPISSFSSTSVIWPETPTQPPVLTGATGTSPAPTTFGS497
Qy	399	-----NLAOMSP--TFMTTLKTSQPLMTPTGTLSSPASLVTCPIAVOTTAGKQLSLTHP451
Db	498	TDSSTSLHLTPSTALSTIVSTSQ--VPISPSTHSSI--LQTP-----STPSSLQTSLTST549
Qy	452	EILVPQISTEGGISTERNRV-----IVDATTGLIPLTSVPTSAKEMTKTLGVTAEYSPA505
Db	550	SEFTTESFTRG--STSTNAILTSESTIWSSTPTIIMSSSPSSA-----SITPVPSTT600
Qy	506	SRSGLTSPSPOTTVYTAEDLAPKSAF-FAVQSTQSQPTLLSSASVNSCAVNPCLHNGE564
Db	601	IHSVPSP-----YIFSTENVGSASITGFPSSLSSATSTSTSSSLTALTITTFPSYI655
Qy	565	CVADNTSRGYHCRCP-----PSWQGDGCSVDVN--ECLGNP-CPSTATCNTQGSF612
Db	656	SLPSTP-----CPGTITIVPASPTDC-VEMDSTSEATSPPTPLTVFPFTE---705
Qy	613	ICKCPVGQLEKICNLVRFTVFTEFKLRTFLNTTVEKHSDLQEVENEI-----661
Db	706	MVTCPTSTSIQ-----TTLTYMDTSMMPESSESI SPNASSSTGTG747
Qy	662	TKTLNMGF-----SALPYSYRS---TVHAS693
Db	748	TVPTNTVFTSLRTPSETWLSNSSVIPLEPLPGVSTIPLTMKPSSSLPTILRTSKSTHPS807
Qy	684	---RESNAVVISLQTTFSLASN-----VTLFLADRMQKCVN-----SKS721
Db	808	PPTRTSETPVATQTPTTLTSRRTRITSQMTTQSTLTTLTTAGTCNGGTWEOGCACLP867
Qy	722	--SAEVCQLLGSRRIFRAGSLCKRKSPCEDKDTISCTDLDGVALCOCKSGYGFQFNKMDH779
Db	868	GFSGDRCOL---QTRCQNG-----QWDGLK-CQCPSTFY-----GS900
Qy	780	SCRACEDGYRLENET-CMSCPFGGLGNCGNP810
Db	901	SCEFAVEQVDLDAEDFCRHAGLHLQG--CGDP930
RESULT	7	
Q9UKW9	90UKW9	PRELIMINARY; PRT; 1217 AA.
AC	Q9UKW9;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
DE	INTESTINAL MUCIN 3 (FRAGMENT).	
GN	MUC3.	
OS	Homo sapiens (Human).	

Db	759	RLPRTSETWLNSSVILPLPGCVSTIPLTMPKSSSLPILRTSKSTHPSPPTTRTSETP-	817
Qy	526	LAPKSATFAVQSSTQSPTTLSS-----SASVNCVAVNCLHNGECVADNTSRG	573
Db	818	-----VATTQTPTLTLSRTRTBITSQMTTSTLTITA-----GTCDNGGTWEQ	860
Qy	574	YHCRCPFSWGDDCSVDVNECLSNPCPSTATCNN--TQGSFICKPVGQLEKIGICNLVR	631
Db	861	GQCACLPFGSGDRCOLQTR-----CQNGGQMDGLKCCQP-----	894
Qy	632	TFVTFEFLKRTFLNTTVE--KHSDLOEVENETKTLNMCFSALPSVIRSTVHASRESNA	688
Db	895	-----STFYGSCEFAVEQVDLWTEVGEVMSVDQQFSPLNDNTSQAYRDFNK	945
Qy	689	V-----VISLQTTFSLASNVTFD-----LAD	710
Db	946	TFWNQMKIIFADMGQFTFKGVEILSLRNGSIVDYLVLLEMPSPQLESEYEQVKTTLKE	1000
Qy	711	RMQKRC--VNSKSSSAEVC-----QLLGSQRRIFRAGSLCKRKSPECDKDTSI----	757
Db	1006	GLQNASQDVNSCQDSQTLCKFKPDSIKVNNNSKTELTPAAICRAAPTGYEEFVPLVEAT	106
Qy	758	DLDGVALC-----OCKSGYGFQFNKMDHISCRACEDGYRLNENETCMSCPPG----	805
Db	1066	RLRCVTKCTSGVDNAIDCHQGCQCVELTSGTTCRCYSTDTHWFSGRCEVAVHWRALVGGL	112
Qy	806	NCGNPYQLITVVIAAAGGGLLLIIGIALIVTCRC	839
Db	1126	-----TAGAALLVLLLLALGVRAVR	1145

```

RESULT      8
ID          Q9C105 PRELIMINARY; PRT; 1236 AA.
AC          Q9C105;
DT          01-JUN-2001 (TEMBLrel. 17, Created)
DT          01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT          01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE          PUTATIVE GLUCOAMYLASE I (ALPHA-1,4-GLUCAN GLUCOSIDASE), EXTRACELLULAR
DE          STARCH-DegrADING ENZYME, BY SIMILARITY TO S. CEREVISIAE STAI, CONTAIN
DE          CHITININASE FAMILY SIGNATURE.
GN          SPAPBIE7.04C.
OS          Schizosaccharomyces pombe (Fission yeast).
OC          Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC          Schizosaccharomycetales; Schizosaccharomycetaceae;
OC          Schizosaccharomycetes.
ON          NCBI_TaxID=4896;
RX          [1]
RN          SEQUENCE FROM N.A.
RC          STRAIN=972H-;
RA          Wood V., RayJandream M.A., Barrell B.G., Seeger K., Harris D.;
RL          Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR          EMBL; AL590605; CAC36921.1; -.
DR          InterPro; IPRO01917; AminoTransf_2.
DR          InterPro; IPRO01579; Chitinase_2.
DR          Pfam; PF00192; chitinase_2; 2.
DR          PROSITE; PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
SQ          SEQUENCE 1236 AA; 123387 MW; 5A2D33A30B87CDD8 CRC64;

Query Match       7.3%; Score 352.5; DB 3; Length 1236;
Best Local Similarity 24.6%; Pred. No. 3.3e-13;
Matches 206; Conservative 132; Mismatches 361; Indels 139; Gaps

Qy    26 TNSTTSDVGTGSSASYPEGVNASVLTFQ-----SDSTVQGSGSHLTALGRSY----- 72
      | : : : | : ||| | : | : | : | : | : | : | : | : | : | : | : | : |
Db    378 TVTSVSDEPTITASGATSVTTTKTDFTVTIVTSTFLIASDSSTSIIVSSYSTVT 437
      | : : : | : ||| | : | : | : | : | : | : | : | : | : | : | : | : |

Qy    73 ---SESSSTSSSELNSAPRGERSIAGISGYQVRGTAEQRFTSSDHHTD----- 119
      | : : : | : ||| | : | : | : | : | : | : | : | : | : | : | : | : |
Db    438 QPASTRVQTIVTVSSISTSVKQPATASVASSVSPSSVPQSSTPIISSSSSASPQSL 497
      | : : : | : ||| | : | : | : | : | : | : | : | : | : | : | : | : |

ov    120 -----TYLSSTFTFKERALLSITDNSSSSDIVESTSYIKITSNHSHSVSSFSHAOTE 172

```

```
Db 498 STSEVSEVSTLLSSAIPSPSSSSPSSSISSPMTSV-LSSSSSIPTSSSDSSS 556
Qy 173 RSNISSYGEYAQSPESPVLHTNLPSYPTINNPNTSVVLDTAEFVSDSSSSSSS 232
Db 557 ITTSSGSSSIPTSSSV--SSILSSSTSPSSLSI-----SSSTSTSTFS 604
Qy 233 SSSSGPPLPLPSVQSQHLLFESSILPSTRASVHLKST--SDATPSSSSPSPVSLTT 290
Db 605 SASTSSPSSSISSSSSTILSSPTPT-SSLMTSSSIISGSSIISSSTIPSSSL 663
Qy 291 ST-----SAPLSVQOTLPQSSSTPVLPRARETPVTFQSTMT-SEMTM 334
Db 664 STYSSVIPSSSTLVSSSSSLVSSPVASSSSPI-----PSSSLVSTYASLSNI 716
Qy 335 LHS--SOTADLKOSTP---HQEKVIESKPSLVSLPTSTKAVTT-----NSPLPPS 383
Db 717 THSLSLTAMSSSAIPTSVNSSTLITASSNTLLSSITSSSAIVSVSTVNISSNLP 776
Qy 384 -----LTSSTEQTLPATSTNLAQMSPTFTTILKTS--QPLMTPTGTLSSLASLVG 434
Db 777 TASSQQLNTSLATSLYLSSSSSRTISTSTNEYNTSFHAPTSSSTSSSTSLAAN 836
Qy 435 PIAYQTAGKOLSTHPEILVPOISTEGGISTERNRIVDATTGLI---PLTSVPTSAKE 491
Db 837 K-GVNSNSITSLNLE-----STSSVTSTAYTTDSVTSTALTSGQPSVVSSLS 886
Qy 492 MTKLG-----VTAYSPASRSLGSPQTTVVSTAEDLAPKATFAVQSTQ----- 540
Db 887 STTSLSTIPVTSVAPAVTSTGETS---SVGSGTDSATSSWTAEITSSAITSSVAA 943
Qy 541 --SPTTLSSASVNSCA-VNPLCHNGECVADNTSRGYHCRCPPMQGDGCDVDVNECLSN 597
Db 944 SVTFTSSSSASSWSSEVDP-----STAASATGSSSTSSATASVSGSSSTSVATASATD 998
Qy 598 PCPSTATCNGTQSFICKCPVGYOLEKICNLVRFVTEFKRKRTFLNTVEKHSDLQEV 657
Db 999 SSTSSIAAASVTGSSSVATASVATDSSSTSVATASATD-----SSTSSIAVASVTGS 1051
Qy 658 ENETKLNCFALPSVIRSTVHASRESNAVVISLQTTFLASNLVFLDLADMOKCVN 717
Db 1052 STSSVATASATDSSSVATASVATISGSSSIATASV--TGSPTSSVAVSTSSVEGTAS 1109
Qy 718 SCKSAEVCQLGSGRRIFRAGSLCKRKSPECDKDTICTDLGVALCQCKSGYGFQFN 775
Db 1110 STIAAASATLSSDA---ASGSSTVSSATSSSSNAATTADSSVTDTDPNSDNFAN 1164

RESULT 9
Q9N4S7
ID Q9N4S7 PRELIMINARY; PRT; 1079 AA.
AC Q9N4S7
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Y51B11A.1 PROTEIN.
GN Y51B11A.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Cotton M.;
```

```
RT "The sequence of C. elegans cosmid Y51B11A.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006797; AF60743.1; -.
DR InterPro; IPR002965; P_Rich_extensn.
DR PRINTS; PRO1217; PRICHEXTENS.
SQ SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;
```

```
Query Match 7.28; Score 348; DB 5; Length 1079;
Best Local Similarity 24.5%; Pred. No. 5.3e-13;
Matches 176; Conservative 102; Mismatches 324; Indels 116; Gaps 26;

Qy 2 SOTETVSRSVAPMGGEGITAHWLL-----TNSTTSDVTGSSASYPEGVNASVLTFQSDS 56
Db 291 TSTPEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITA--PETTSTPEPSS--STT 347
Qy 57 TVQSGGSHALGDRSYSESSSTSSSESLNSAPR---GERSIAGISYQVVRGTAIEQRTS 113
Db 348 PVQTTTITAPETTTEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTS 407
Qy 114 SDHTDHT---VLSSTFKGERALLSITDNSSSDIVESTSYIKISNSHSEYSSFSHAQ 170
Db 408 TEPPSSSTTPVQTTTITAPETT--STEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQ 465
Qy 171 TERNSISYDGEYAQSPESPVLHTSNLPSYPTINNPNTSVV-----LDTDAEFVSDSS 225
Db 466 TTTTITAPETT-----TTSTPEP--SSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITA 517
Qy 226 SSSSSSSSSSSSGGPL-----PLPSVSQSHLFFSLIPSTRASVHLKSTSDASTPWSSS 280
Db 518 PETTSTPEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTSTPEPSSSNT 575
Qy 281 PSPLPVSLTT-----STAPLSVSOTILPO-----SSSTPVL----- 312
Db 576 TTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPET 635
Qy 313 -----PRARETPVTSFQSTMTSFMTLHSSQADLKSQTHQEKVITESKPSLVSLP 367
Db 636 TSTPEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTSTPEPSSSNT 695
Qy 368 TESTKAVTNSPLPSLATES---STEQLPATSN---LAOMSPFTTIL----- 412
Db 696 TTTTITAPETTSTPEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTSTEP 755
Qy 413 --KTSQPLMTTPGTLSSSTASL-----VTGPIAVOTTACKQLSLTHPEILVPOISTEGGIS 465
Db 756 PSSSTTPVQTTTITAPETTSTPEPSSSNTTPVQTTTITAPETTSTEP-----PSSSTS---P 808
Qy 466 TERNRIVDATATGLIPTSVPTSAREMTKLGVTAEYSPASRSLGTSPSPQTTVVSTAED 525
Db 809 VQTTTITAPETTSTPEPSSSNTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTST 868
Qy 526 LAPASATFAVQSTT-QSPTTLSSASVNSCAVNCPLHNGECVADNTSRGYHCRCPSPSQG 584
Db 869 EPPSSSTTPVQTTTITAPETTSTPEP--SSSTTPVQTTTITAPETTSTPE-----PPSSST 921
Qy 585 ---DDCSVDVNECLSNPCPSTAT--CNNTQGSFCKCPVGYOLEKICNLVRTFTVFTEF 637
Db 922 TPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETT 974
```

```
RESULT 10
Q29071
ID Q29071 PRELIMINARY; PRT; 528 AA.
AC Q29071;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
```

DE GASTRIC MUCIN (FRAGMENT).
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX MEDLINE=95275264; PubMed=7755593;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA LaMont J.T.;
RT "Isolation and characterization of cDNA clones encoding pig gastric
RT mucin.";
RL Biochem. J. 308:89-96(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX MEDLINE=94102478; PubMed=7506218;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA LaMont J.T.;
RT "pig gastric mucin: isolation and characterization of a cDNA clone
RT with a novel tandem repeat";
RL Gastroenterology 106:200-200(1994).
DR EMBL; U10281; AAC48526.1; -.
FT NON_TER 1
FT NON_TER 528
SQ SEQUENCE 528 AA; 49907 MW; 0BF0F6879203B2EA CRC64;

Query Match 7.1%, Score 347; DB 6; Length 528;
Best Local Similarity 27.9%; Pred. No. 2.5e-13;
Matches 161; Conservative 81; Mismatches 204; Indels 132; Gaps 23;

QY 10 SVAPMRGEITAHWLLNSTSDADVTGSSASYPGYNASVLTQFSDS-----TWVSGG 62
DB 35 SVQSSSGSAP-----TTSATVQ-TSSSSPPISSTISVOTSSSSSVPTTTSVQSS 88
QY 63 SHTALGRSYSESSSTSSSLNSAPRGRSISAGYQVGRGTAIBQRTSSDHTDTYL 122
DB 89 SSSAPTTRATSVQSSSSSAPISSTT-----SVQSSSGSVPTTSATVQSSSSSAPT 143
QY 123 SSTTKGERALLSTDNSSSDIVESTSYIKINSNHSHEYSFSHAQTERSNISSVDGE 182
DB 144 SAT-----SVQSSSSSPISSTSVQSSSSSAPTTS----- 179
QY 183 YAQPSTES--PVLHTSNLPVTPIINPNTSVLWLDTAERFVSDSSSS-----SSSSSSS 236
DB 180 -VQPSSSSPPI-----SSTSVQTSSSSVPTTTSVQSSSSSVPTTSATSVRSS 231
QY 237 SGPLPLPSVQSQHLFSSILPSTRASVHLKSTSDASTPSSSPPLP-----VSLTST 292
DB 232 SSSSTPIPST-----TSVQSSSSSAPTTSATS--VQPSSSSTPIPSTTSVQPSSSS 282
QY 293 SAPLSVQTLTPOSSSTPVLPRARETPTVTFQRTSTWTFMTLHSSQTADLKSGSTPHQE 352
DB 283 SAPTTSATSVQPSSSSP-----BIS-----TISVQPSSSSSSP 317
QY 353 KVTESKPSLV-SLPTESTKAVTNSPLPLPSTESTEQTLPATSNLAQMSPFTTTI 411
DB 318 TTSTTSVQPSSSGAPTTSATSVQPSSSSPPI--SSTISVQPSSSS-----SPTTSTTS 371
QY 412 LKTSQPLMTTPTGLTSSPASLVGTPIAVQTTAGKQLSLTHPEILVPOISTEGGISTERNRY 471
DB 372 VQPS-----SSGAPTTSATSVQPSSSS-----VPTTSATSVRSSSSST 412
QY 472 IVDATTCGLIP--LTSVPT-----SAKEMTKLGVTAEYSPASRLGCTSPSQTTVVST 522
DB 413 PIPPTTSVQPSSSSVPTTSATSVQTSSSSTPIPSTTSVQPSSSSAPTTS-ATSQPS 471
QY 523 AEDLAPKATFAVQ--SSTQSPSTLSSASVNSCAVNP 558
DB 472 SSSSPPISSISVQPSSSSSSPTTSTTSVQPSSSGAP 509

RESULT 11
P97881
ID P97881 PRELIMINARY; PRT; 547 AA.
AC P97881;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE PUTATIVE CELL SURFACE ANTIGEN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RA Hajdu A., Flanagan P.R.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U89744; AAB49894.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_2.
DR PROSITE; PS00024; SEA; 1.
SQ SEQUENCE 547 AA; 57650 MW; EA86608C704080CF CRC64;

Query Match 7.1%, Score 343.5; DB 11; Length 547;
Best Local Similarity 24.4%; Pred. No. 4.3e-13;
Matches 160; Conservative 66; Mismatches 198; Indels 233; Gaps 31;

QY 268 KSTSDASTPSSSPPLVSLTSTAPLSVSQTTLQSSSTPVLPRARETPVTSFQTST 327
DB 3 QSSGGTSTPTTATQP---TSTSTQP---GTTQLLSTSTPTTTATQPT-----STST 50
QY 328 MTSFMTMLHSQSOTADLKQSQTPHQEVITESKPSLSLSP-TESTKAVTNSPLPPLS 386
DB 51 QTPCTTQLPSTTSPPTTATQP---TSTSTQP---TSTSTQP---TSTSTQP---TST 106
QY 387 SSTQTLPATSTNLAQMSPTFTTILKTSQPLMT---TPGTLSTASLVGTPIAVQTAG 443
DB 107 PGTTLQPSSTST-----PTTAT-----OPTSTASQTPGT----- 136
QY 444 KQLSLTHPEILVPOISTEGGISTERNRVIVDATTGLTSTVPTSAKEMTKLGVTAEYS 503
DB 137 -----TQP-----PGGASS-----PTTIV-----TQPTGSSSQTP 161
QY 504 PASRLGTSPTSPQTTVVSTAEADLAPKATFAVQSSSTQSPSTTLSSSASVNSCAVNPCLHNG 563
DB 162 GTTQPPGGASTPTTIVTQ-----PTGSSSSTSTQTPPGGASSST----- 201
QY 564 ECVADNTSRGHCRCPPSWGGDDCSVDVNECLSNPCPSTATCNNTQSGFICKCPVGQL- 622
DB 202 --VTSSSTSGN-----DPCNSNPCKSPASCVKLYDSYFCLEGYYN 243
QY 623 EKGTCNLVRFVTVEFKLRTFLNTTVEKHSQDLP-----VENEITKT-- 664
DB 244 NSSSCVGTTPGE-----IGMSVNETTDLKNSVNYQTLHSSVVKFFENTFKKTDY 296
QY 665 -----LNMCFSAIPSYIRSTVHASRESNNAVVISLQTTF-----SLAS----- 701
DB 297 GQTVILKVSXKDSLMS-SRSMRAATQT--VVVSVNVNFGENTKEDESVASVIKEAVKTD 353
QY 702 -NVTFLDLADBMQ--KCVNSCKSAEVCQLLGSORRIFRAGSLCKRK----- 745
DB 354 NNVERYEQDRCDYIGCV--KSGSNVC-----RNLQCTCKPGLERLNPQVPFC 400
QY 746 -SPECCKDTS-----ICTDLGDVALCOCKSGYFQFNKMDHSCRACEDGYRLENETCMSCP 799
DB 401 VAPTCSEPCSAEKKQJCLKKDNGAM-EC-----CCMAGYRKANGKCECP 444

```
Qy 800 FGLGLGNGPNYQLITVVIAAAGGLLILGIALIVTCRKNKNDISKLIKSGDFQ 856
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 445 FGYSMDCKDQFOLLITVGTIAGAFILILLIVFVSMRSNKK-----KSGEQ 494

RESULT 12
Q21027 PRELIMINARY; PRT; 786 AA.
AC Q21027;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE F59A6.3 PROTEIN.
GN F59A6.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium. ";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Waterston R.;
RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U41994; AAK31523.1; -
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR SMART; SM00032; CCP; 1.
SQ SEQUENCE 786 AA; 79211 MW; 60425E32F083B3CB CRC64;

Query Match 7.0%; Score 339; DB 5; Length 786;
Best Local Similarity 24.0%; Pred. No. 1.3e-12;
Matches 180; Conservative 105; Mismatches 334; Indels 130; Gaps 25;

Qy 26 TNSTTSADVTCSSAYPEGVNASVLTFSDTVSGGSHLTALGDRSYSESSSTSSSESLN 85
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 107 TTSTASTFTISSSLKSTSTSDSTSTPRISTTTDTKDTTDPVSDQSSSTSPHETTR 166
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 86 SSAPRGERSIAGIS-YGOVRGTAIEQRTS-----SDHTDHTYLSFTFKGERALLSIT 137
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 167 DTTTEGTSEDSITYGTERSSSPKPTSEFSTESDFTSTSETS-----SIE 219
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 138 DNSSSDSDIVESSTYKISNSHSEYFSFSA-----QTERSNISSYDGEYAQPSTSPV 192
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 TNSSTSPV--STPEYDSTSGNSETTESDGTITVTTKDDTSTVSGDNSGSSSTSEF 277
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 193 LHTNLPSTYPTINMPTNSVVLDTDAEFVDSSSSSSSSSSS--SSGPPPLPSPVQSHH 251
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 278 KNTETTTGPGTVSEPSSESSDLSSSVSDRSTDSQDRTEIGLQGPIL----- 327
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 252 LFSSILPSTRASVHLLKSTSDASTPWSSSPPLPVSLTSTASPLSVQTLPOSSSTPV 311
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 -----SDSNPNPDEFTTSALTSGSTSTTSRASSADDPTTTPGST-- 367
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 312 LPRAREPVTSTFOSTWTS---FMTMLHSSQADLKSQSTPHQEKVITESKPSILV-LP 367
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 -----TSGSTASTSGSLFSTSLGSSQSPGSSVSTTP-----GPSTISGIS 408
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 368 TESTKAVTNSPLPPSLTESTEOTLPATNTNLQMSPTFTTTLKTSQPLMTTPG-TLS 426
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 409 QSTTSGPTTTS--EPSTTSGTVDSTSGPSTT--SGPSTTLGTQTSTTSGP--STTPGSTIT 464
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 427 ST--ASLVTPG-----IAVQTAGKQLSLTHPEILVPOISTEGGISTERNRIVDAITGL 479
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 465 TSSASTTSGPSTSGSTVSTSGQSTSGTKSTTSGPTTSSGSPSTVSERT----- 516
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 480 IPLTSVPTSAKEMTTKLGVTAEYSPASRLGTSPSPQTTVVSTABDLAPKATPAVQSSST 539
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 517 LSTTSGPSTTSGPSTTSGSTVSTTPGAST--TSGSTQSTTSGPSTSSGSPSTASRSTVSTT 574
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 540 QSPTTSL-----SSASVNCVAVNCPCLHNGECVADNTRSGYHCRPPSQWGDGCSV 589
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 575 SGPSTTSGPSTTSGPSTTSGTKSTTSGPSTTSGKNISTVSGK-----LTGSTTSA 625
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 590 DVNECLSNPCSTATCANTQGSFICKCPGVQLEKIGICNLVTRTFVTERKLRKTFNTTIVE 649
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 626 TISSAFGNVFTTKSPSSNGGTTSSGKNFSQNTTSAANGTTQAYNNGKSGSTLPTNSS 685
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 650 KHSDLQEVENEITKTLNMFSAIPSYIRSTVHASRESNAVVISLQTTFSLASNVF----- 704
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 686 GSSDSTSPSTFTVSYNNTNSI--TCIEPSNLNLTYSSTSOIKSSYVGLITQVCPQ 744
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 705 --LFDLA---DRMOKCVNSCK--SSAECV 726
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 745 SYVEIALQPVKIYKCLATGWAGSPENC 773
||| : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
Q9H195 PRELIMINARY; PRT; 901 AA.
AC Q9H195;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MUC3B MUCIN (FRAGMENT).
GN MUC3B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pratt W.S., Crawley S., Hicks J., Ho J., Nash M., Kim Y.S., Gum J.R.,
RA Swallow D.M.;
RT "Multiple transcripts of MUC3: Evidence for two genes, MUC3A and
RT MUC3B. ";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ291390; CAC19572.1; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER 1
SQ SEQUENCE 901 AA; 96192 MW; F98FC817494ECD99 CRC64;

Query Match 6.9%; Score 337.5; DB 4; Length 901;
Best Local Similarity 22.0%; Pred. No. 1.8e-12;
Matches 211; Conservative 117; Mismatches 289; Indels 341; Gaps 44;

Qy 132 ALLSITDNSSSDIVESSTYKISNSHSEYFSFSHAQTERSNISSYDGEYAQPSTESP 191
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 SIITETTSHTSPFSSTIHSSTVSSSTTATSPPTTAETGVTSTPS-----SPSSLST 59
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 192 VLHSTNLPSTYPTINMPTNSVVLDTDAEFVDSSSSSSSSSSSGPPPLPSPVQSHH 251
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 DIPTSLRLTLPISLSTSTSTTTTDLPSIPTDLSLPIHIISSPSIOSTETSSLVG 119
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 252 LFSSILPSTRASVHLLKSTSDAS-----TPWSSSP-----SPLPVSIL- 289
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 TTSPTMTSVTRATLSTENTPISSFSTSIIVTPTPTTQAPPVLMASATGQTSPVPTTTF 179
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 290 -----TSTAPLSVSQTLTPOSSS-----TPVLPRARETPTVSF----- 323
Db 180 GMSDSSSTLHLLTPSALSKIMTSOPPISTHSSLTQTPSIP-SLQSLSTSTSEFT 238
QY 324 -----QTST---MTSFMTLMHSQTADLKSQS-----TPHQKVI-----TE 357
Db 239 ESFTRGSTNAILTSTIWSSTPTIIMSSSPSSASITPVFATTIHSVPSPPIFSTE 298
QY 358 S-KPSLSVLPTSTKATVNTSPLPPLSTESSTEQTLPAITSNLAQSP---TFTTIL- 412
Db 299 NVGSASITAFPSLSSTSTSTPSSTSLTALTET-PSFYISLPSPTPCPGTTIIVP 357
QY 413 -----KTSQSLMTTPGTL-SSTASLVTPG--IAVQTAGKQL-----SLTHP 451
Db 358 ASPTDPCVEMDPSTEATSP--TTPLVFFPFTTEWTCPSISNQTLATHMDTSSMT-P 414
QY 452 E---ILLVPQISTEGIST-----ERNRIVDATGTL--IPLTSVPT 487
Db 415 ESESIIPNASSSTGTGTPTNTVFTSTRLPTSETWLSNNSVIPPTPLPGVSTIPLTKPS 474
QY 488 SAKEMTKLGVTAIYSFASBSLGTSPSPQTVVSTAEADLAPKSFATFVQSTQSTPTLSS 547
Db 475 SS---LPTIL-----RTSSKTHPSPPTARTSQTS-----VATQTPPTLT- 512
QY 548 SASVNSCAVNPCLH-----NGECVADNTSRGYHCRCPSPSQGDCCSDVDNCELS 596
Db 513 -----TCRTPTITSMMTTQSTLTATTACTDNGGTWEOGQCACLPFGSGDCOLQTR---- 563
QY 597 NPCSTATCN--TOGSFICKPVGYLEKIGICNLVFTFTEFKLTKTFLNTVE---KH 651
Db 564 -----CONGGOWDLKQCP-----STFYGSCEFAVEQ 592
QY 652 SDLOVEVNETIKTNMCFALSALPVSIRVTHASRESNAV-----V 690
Db 593 VLDLVTEVGMESVVDQFQSPDLNDNTSQAYRDNKTFNQMKIFADWQGFYKGVFI 652
QY 691 ISLQTTESLASNVFLD-----LADMOKC-----VNSKSSSAEVCOLLG 730
Db 653 LSLRNGSTIVVDYLVLEMPFSPQLESEYEQVKTTLKEGLQNASQDANSQDQALC----- 708
QY 731 SQRIFRAGSLCKRKSPECDKDSICTDLQVALCO--CKSGYFQF----- 774
Db 709 -----FKPDSIKVNNNSK-----TELTPAICRAAPTGYEEFYPLVATRLRCVT 755
QY 775 -----NKMDSHCRACEDGYRLENETCMSCPFGLGLINC-----GNPYQLITVV 817
Db 756 KCTSGLDNAID-----CHQG-----OCVLETSPACRCYSTDTHWFGSPCEVTVH 801
QY 818 IAAAGGGLLLILGIALIVTCRKNKNDISKLIFKSGDFQMSPYAEYKPNRPSQEWGRE 875
Db 802 WRALVGG--LTAGAALLVL-----LLALG-----VRAVRSEWNGRQ 836

RESULT 14
Q939N5 PRELIMINARY; PRT; 3072 AA.
ID AC Q939N5
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE PLATELET BINDING PROTEIN GSPB.
GN GSPB.
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M99;
RA Bensing B.A., Sullam P.M.;
RT "An accessory sec locus of Streptococcus gordonii is required for
RT export of GspB and for platelet binding.";
```

```
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY028381; AAL13053.1;
SQ SEQUENCE 3072 AA; 285770 MW; 0BJ48372697CF7F2 CRC64;

Query Match 6.9%; Score 336; DB 2; Length 3072;
Best Local Similarity 24.7%; Pred. No. 1.1e-11;
Matches 140; Conservative 112; Mismatches 281; Indels 34; Gaps 10;

QY 1 MSQTEVSRSVAPMRGGEITAHLLTNTSTTSADVTS-----SASYEGVNASVL 51
Db 637 ISSASVSASTSISTGSGVSAS---ESASTSSVSASESASTSASVSASESASTS 693
QY 52 QFSQSTVQSGSHALGDRSYSESSSTSSSLNSLSPRGERSIAGISYGVGRGTAEQR 111
Db 694 STASTASVSASTS-ASTSASTSASKASTSASVSASTSASTSASVSASESASTS 752
QY 112 TSDHTOHTYLSSTFTTKGERALLSITDNSSSDIVESSTSYKINSNHSYSFS---H 168
Db 753 ASTSASTSASVSASTSASTSASVSASESASTSASVSASTSASTSASVSASESAST 812
QY 169 AQTERSNISSVDGEYAQPSTESPVLHTSNLPSYPTINMPTSVVLDTDAEFVSDSS 228
Db 813 ASTSASTSASVSAS-ASASTSASVSASTS-----ASTSASVSASAS--ASTS 859
QY 229 SSSSSSSSGPPLPLPSVSQSHHLFSILPSTRASVHLLKSTSDASTPWSSSPPLPVSL 288
Db 860 SASVSASTSASVSASESASTSASVSASESASTSASVSASESASTSASVSASESAST 919
QY 289 TTSAPLSVSQTLTPOSSSTPVLPRAETPVTSFQSTMTSFTMTLHSSQTADLKOST 348
Db 920 SASTSASTSAS-VSASESASTSASVSASESASTSASVSASESAST---SASVASE 975
QY 349 PHQEKVITESKPSLIVSLPTSTKAVTNSPLPPSLTESSTEQTLPATSTNLAQSPFT 408
Db 976 SASVSASTSASTSASVSASTSASTSASVSASTSASTSASVSASESASTSASVASE 1035
QY 409 TTILKTSQPLMTPTGLSTASLTGPIAVQTAGKQLSLTHPEILVPOISTEGGISTER 468
Db 1036 SASVSASESASTSASVSASTSASTSASVSASESASTSASVSASESASTSAS-ES 1094
QY 469 NRVIYDATTGLPILTSVPTSAKEMTKLGVTAIYSFASRSRGTSPTTIVVTAEDLAP 528
Db 1095 TSASVSASESASTSASVSASESSTSSASVSASESSTSSASVSASESASTSASV 1154
QY 529 KSATFVQSSSQSPSTLSSASVNSCA 555
Db 1155 TSASESASESASTSASVSASESASTSA 1181

RESULT 15
Q9KWR3 PRELIMINARY; PRT; 2178 AA.
ID AC Q9KWR3
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE STREPTOCOCCAL HEMAGGLUTININ.
GN HSA.
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DLL;
RA Takahashi Y., Konishi K., Yoshikawa M.;
RT "Cloning and characterization of the gene encoding a hemagglutinin of
RT Streptococcus gordonii DLI.";
RL Streptococcus gordonii DLI.
Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB029393; BA97453.1;
DR InterPro; IPR004089; Chemotaxis_transducer.
DR InterPro; IPR001899; Gram_pos_anchor.
```

DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 2178 AA; 203507 MW; 75CC27E27F41DA8C CRC64;

Query Match 6.9%; Score 334.5; DB 2; Length 2178;
Best Local Similarity 22.0%; Pred. No. 8.5e-12;
Matches 177; Conservative 145; Mismatches 367; Indels 115; Gaps 19;

QY 2 SOTETVSRVAPMRGGTEITAHLLTNS-TTSDADVTGS-----SASYPEGYNASVLT 51
Db 1434 SESAYTSASVASESGTSASVSASESTASVSASESTASVSASESTASVSASESTASVS 1493
QY 52 QFSDST---VQSGSHALGDRSYSESSSTSSSESLNS-----SA 88
Db 1494 SASSTASVSASGASTASVSASESTASVSASTASVSASESTASVSASESTASVS 1553
QY 89 PRGRSTAGISYQGVRCGTAEQRTSSDHTDHTYLSSTFTTGERALLITDNSSSDIV-- 146
Db 1554 SESASTASVSASESTASVSSTASVSASESTASVSASESTASVSASESTASVS 1613
QY 147 -ESSTSVIKISNSSHSYSEFSHAQTERSNISSYDGEYAQPSTESPVLHTSNLPSYPTI 205
Db 1614 SESSTASVS-ASESTASVSASESTASVSASESTASVSASESTASVSASESTASVS 1672
QY 206 NMPNTSYVLDTADEFVSDSSSSSSSSSSSGPPLPLPSVQSHHLFSILPSTRASVH 265
Db 1673 ASESASTASVSASESASSASVSASKASMSASVLAASEASTASVSASESTASVS 1732
QY 266 LLKSTDASTPMSSPPLPVSLTTSAPLSVQTLPOSSSTPVLPRARETPVTSFQT 325
Db 1733 ASESASTASVSASESTASVSASESTASVS-VSASESTASVSASESTASVS 1791
QY 326 STWTSFMTLHSSQTADLKQSOTPHQKVITESKPSLVSLPTSTKAVTNSLPLPSLT 385
Db 1792 SASTS-----ASTSASVSASESTASVSASESTASVSASESTASVSAS 1842
QY 386 ESSTEQTLPATSTLQAQMSPTFTTILKTQPLMTTPGTLSTASLVGTPIAVOTTAGKQ 445
Db 1843 ESVSASESTASVSASTASVSASTASVSSESTASVSASESTASVSASESTASVS 1902
QY 446 LSLTHPE--ILVPOISTEGGISTERNVIVDATGTLPLTSVPTSAKEMTKLGVTAEYS 503
Db 1903 ASVSASESTASISASESSSTAS---VSASESTASVSASTSTSTASVSASESA 1959
QY 504 PASHSLGTSPPQTVVYTAEDLAPKATFAVQSTQSTPTLSSASVNSCAVNPCLHNG 563
Db 1960 STSASVFASESTASVSASESTASVSASTASVSASTASVSASESTAS 2011
QY 564 ECVADNTRSGYHCRPPSWQDDCSVDVNECLSNPCP-STATCNNTQGSFICKCPVGYL 622
Db 2012 -----SISASESTASISASESSSTASVSASTASVS 2047
QY 623 EKGICNLVRFVTEFKLRTFTNTVTEKHSDLQEVENEITKTLNMFSPALPSYIRSTVHA 682
Db 2048 SVSASESTSTSVISASESVISSTVSQSMVSE-----SLSLSVSTSTLHSQLNG 2098
QY 683 SRESNAVVISLQTTFLASNV--TLFD-----LADRMOK-----CVNSCKSSAEVCQ 727
Db 2099 IYSELNSLSLSLSMSQSLSQSLSQSTSATQSMHDIRISKQLPRTGESEKASILA 2158
QY 728 L-LGSRRIFRAGSLCKRKSPED 750
Db 2159 LGTGLGLAFK-----KRKNESE 2177

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 20:05:41 ; Search time 955.43 Seconds
(without alignments)
12492.793 Million cell updates/sec

Title: US-09-840-746-2
Perfect score: 6952
Sequence: 1 gttcgatgaagaattgccg.....tactgggaagctatagggtg 6952

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Result No.	Score	Query Match	Length	ID	Description
1	5098.4	73.3	6407	20 AAX26246	CDNA sequence of r
2	5098.4	73.3	6407	20 AAV81825	Human rchd528 enco
3	5098.4	73.3	6407	21 AAA88578	Human rchd528 cDNA
4	5098.4	73.3	6407	21 AAZ89796	Human cardiovascular
5	5098.4	73.3	6407	21 AAZ88007	Rchd528 gene seque
6	5042.4	72.5	6407	17 AAT36035	rchd528 gene diffe
7	3952.8	56.9	3975	22 AAK94374	Human full-length
8	3092.8	44.5	3116	22 AAS02424	Human secreted pro
9	3080	44.3	3170	21 AAF18333	Lung cancer associ

SUMMARIES

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

10	831	12.0	3075	22 AAL26437	Human breast cancer
11	708.2	10.2	751	22 AAK92040	Human CDNA 5'-end
12	708.2	10.2	751	22 AAK93703	Human CDNA clone r
13	703	10.1	869	21 AAH31088	Human colon cancer
14	500.8	7.2	537	22 AAL34960	Human musculoskele
15	480	6.9	546	22 AAK92794	Human CDNA 3'-end
16	471	6.8	616	22 ABA06457	Human CDNA SEQ ID
17	471	6.8	616	22 AAS28946	CDNA encoding for
18	471	6.8	616	22 AAS26918	Human CDNA encoding
19	471	6.8	616	22 AAS31583	CDNA encoding nove
20	470.4	6.8	584	21 AAH30808	Human colon cancer
21	386.4	5.6	400	21 AAH30808	Human colon cancer
22	385.8	5.5	400	22 AAS37279	Novel human diagno
23	357	5.1	391	21 AAH30820	Human colon cancer
24	351.4	5.1	781	22 AAS26846	Human CDNA encoding
25	331.2	4.8	351	21 AAA50236	DNA encoding autoa
26	247.2	3.6	252	22 AAL25327	Human breast cancer
27	231	3.3	248	20 AAV88062	EST clone FS173.
28	230.4	3.3	403	22 AAL07581	Human breast cancer
29	191.8	2.8	195	22 AAL16484	Human breast cancer
30	152	2.2	349	22 AAL19125	Human breast cancer
31	150	2.2	343	22 AAL25960	Human breast cancer
32	136	2.0	333	22 AAL08215	Human breast cancer
33	82.6	1.2	344	22 AAK56290	Human immune/haema
34	57.4	0.8	651	23 AAS84956	DNA encoding novel
35	56.8	0.8	824	23 AAS66529	DNA encoding novel
36	56.4	0.8	1072	23 AAS90738	DNA encoding novel
37	56.2	0.8	1050	22 AAD21685	Mutational hot spo
38	56.2	0.8	2849	22 AAD21684	Human retinitis pi
39	56.2	0.8	3489	21 AAA30290	Kaposi's sarcoma-a
40	56.2	0.8	3489	22 AAF82901	Nucleotide sequenc
41	56.2	0.8	32207	20 AAV73805	KSHV LUR DNA (nucl
42	56.2	0.8	137507	19 AAV19941	KSHV long unique c
43	56	0.8	583	22 ABA50512	Human breast cell
44	56	0.8	583	22 ABA68463	Human foetal liver
45	56	0.8	583	22 ABA35452	Probe #13918 for g

ALIGNMENTS

RESULT	1
AAX26246	
ID	AAX26246 standard; CDNA; 6407 BP.
XX	
AC	AAX26246;
XX	
DT	24-MAY-1999 (first entry)
XX	
DE	CDNA sequence of rchd528 gene.
XX	
KW	Fingerprinting gene; rchd502; transmembrane protein; cardiovascular;
KW	fingerprint/target gene; up-regulated; endothelial cell; shear-stress;
KW	atherosclerosis; ischemia; reperfusion; hypertension; restenosis; ds.
XX	
OS	Homo sapiens.
XX	
PN	US5882925-A.
XX	
PD	16-MAR-1999.
XX	
PF	09-FEB-1996; 96US-0599654.
XX	
PR	09-FEB-1996; 96US-0599654.
PR	10-FEB-1995; 95US-0386844.
PR	07-JUN-1995; 95US-0485573.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Falb DA;
XX	
DR	WPI; 1999-214071/18.
DR	P-PSDB; AAW95160.

XX New polynucleotides consisting of residues 1-1929 of the rchd502
PT gene - are differentially expressed in cardiovascular disease
PT states, and can therefore be used to treat and diagnose
PT cardiovascular diseases
XX
PS Disclosure: Fig 30A; 121pp; English.
XX
CC The invention relates to a rchd502 target/fingerprint gene encoding a
CC transmembrane protein. The invention provides cDNAs contained in plasmids
CC pFCUPD502SF (ATCC 69981) and pFCUPD502SJ (ATCC 69982) that encode the
CC rchd502 polypeptide, and are differentially expressed in cardiovascular
CC disease states. Cultured genetically engineered host cell containing the
CC rchd502 polynucleotides in operative association with a nucleotide
CC regulatory element are used for producing a polypeptide rchd502 gene
CC product. Identifying that the fingerprint/target gene rchd502 is
CC differentially expressed (up-regulated) by endothelial cells subjected
CC to shear-stress, provides a tool for the diagnosis and treatment of
CC cardiovascular disease e.g. atherosclerosis, ischemia/reperfusion,
CC hypertension, restenosis. The fingerprint gene is useful for testing the
CC efficacy of candidate drugs in basic research and in clinical trials and
CC or imaging of a diseased cardiovascular tissue. The gene may also be
CC used in screening for ligands of target gene product receptor domains, as
CC well as antagonists of the ligand-receptor interaction.
XX
SQ Sequence 6407 BP; 1594 A; 1762 C; 1520 G; 1509 T; 22 other:

Query Match 73.3%; Score 5098.4; DB 20; Length 6407;
Best Local Similarity 94.3%; Pred. NO. 0;
Matches 5412; Conservative 21; Mismatches 8; Indels 301; Gaps 2;

QY 1 gtcgatgaagaattgcgcgttttcaacaagaagtgaacagcctcgagatgggaac 60
DB 666 gtcgatgaagaattgcgcgttttcaacaagaagtgaacagcctcgagatgggaac 725

QY 61 agagagggcgatgggctgcagaagaatggactgtgcacagccaagagccaccacttc 120
DB 726 agagagggcgatgggctgtcagaagaatggactgtgcacagccaagagccaccacttc 785

QY 121 ggcttgagagcccttcctttctctctctgttgagatggagagactgaccacgctctctag 180
DB 786 ggcttgagagcccttcctttctctctctgttgagatggagagactgaccacgctctctag 845

QY 181 gaagagaaatctctcaggaccagatctctctgtgctgctatttctacagacagcagcttc 240
DB 846 gaagagaaatctctcaggaccagatctctctgtgctgctatttctacagacagcagcttc 905

QY 241 ctctctctcttagacctttctcactctctgaaagtacagagaagcttaacaaactccac 300
DB 906 ctctctctcttagacctttctcactctctgaaagtacagagaagcttaacaaactccac 965

QY 301 tggcctccagagctctcagtcagtcacaacaagaacaaatgcattgtgctaccgtgtccac 360
DB 966 tggcctccagagctctcagtcagtcacaacaagaacaaatgcattgtgctaccgtgtccac 1025

QY 361 tgatggtgcccagagacgctgcgactctttgacggtcagctctggagccctgtgagcaagac 420
DB 1026 tgatggtgcccagagacgctgcgactctttgacggtcagctctggagccctgtgagcaagac 1085

QY 421 agaaggtctcccaaggactccagaattgccagacttcatctcagtcagtcctttctcacc 480
DB 1086 agaaggtctcccaaggactccagaattgccagacttcatctcagtcagtcctttctcacc 1145

QY 481 ctctcagtggaatcagagaagaacagtagagtaactgggaatcccaaggagtagagaatt 540
DB 1146 ctctcagtggaatcagagaagaacagtagagtaactgggaatcccaaggagtagagaatt 1205

QY 541 cattgacctccacagaaaaatgaatttgacttgcgtctcttgcg-tggcaaaaatgattc 599
DB 1206 cattgacctccacagaaaaatgaatttgacttgcgtctcttgcg-tggcaaaaatgattc 1265

QY 600 cccaacctttggagaacatcagcttgccagcagctctgaggtgcacaaatgggaagtcccat 659

DB 1266 cccaacctttggagaacatcagcttgccagcagctctgaggtgcacaaatggagtcacct 1325
QY 660 gtcctcagactgagactgtgtcttaggtcagtcgcacccatgagaggtggagacatcctgc 719
DB 1326 gtcctcagactgagactgtgtcttaggtcagtcgcacccatgagaggtggagacatcctgc 1385
QY 720 acactggctttgaccaaacagcacacatctgcagatgtgacaggaagctctgcttcata 779
DB 1386 acactggctttgaccaaacagcacacatctgcagatgtgacaggaagctctgcttcata 1445
QY 780 tctcgaaggtgtgaatgtctcagttgttgaccagttctcagactctactgtacagttctgg 839
DB 1446 tctcgaaggtgtgaatgtctcagttgttgaccagttctcagactctactgtacagttctgg 1505
QY 840 agaaatgcacacagcattgggagataggattattcagagcttctcattcattcattcttcctc 899
DB 1506 agaaatgcacacagcattgggagataggattattcagagcttctcattcattcattcttcctc 1565
QY 900 ggaagcttgaattcatcagcacacacagctggagaaacttcaa----- 940
DB 1566 ggaagcttgaattcatcagcacacacagctggagaaacttcaaacttggaaagacagccgaga 1625
QY 941 ----- 940

DB 1626 gccaggccaagcactaggtgacagttccgcgaatgcagagagcagacttctgggtgcc 1685
QY 941 ----- 940

DB 1686 ctctctcggcaccacacacttggctactgtcacttgaaacggggaacgcacactgcggtc 1745
QY 941 ----- 940

DB 1746 tgtcacccctcacacacacacagcatgacgacacttcttggggaagcagcgcctgcagc 1805
QY 941 ----- 940

DB 1806 ggccatgcccccagaacacagagggtgctctctgcacgtaaacgtgacggacgacatggg 1865
QY 941 -----tcgctgggattagctacgg 959

DB 1866 cctggtctcagcgtcactcctcccgccctccagctgcaactggagctgctgggattagctacgg 1925

QY 960 tcaagtgcgtggcacagctattgaacaaaggacttccagcgacacacacacacacacta 1019
DB 1926 tcaagtgcgtggcacagctattgaacaaaggacttccagcgacacacacacacacacta 1985

QY 1020 cctgtcactctacttccacaaaggagaacgggcttactgtccattacagataaacagttc 1079
DB 1986 cctgtcactctacttccacaaaggagaacgggcttactgtccattacagataaacagttc 2045

QY 1080 atcctcagacattgtggagagctcaactctcttatattataaaactctcaactcttcacattc 1139
DB 2046 atcctcagacattgtggagagctcaactctcttatattataaaactcttcacattc 2105

QY 1140 agagtattctctctcttctcactgctcagactgagagaagtaaacatctcactatgacgg 1199
DB 2106 agagtattctctctcttctcactgctcagactgagagaagtaaacatctcactatgacgg 2165

QY 1200 ggaatatgctcagccttctcactgagtcgcagttctgtgcatacatccaacacttcgctccta 1259
DB 2166 ggaatatgctcagccttctcactgagtcgcagttctgtgcatacatccaacacttcgctccta 2225

QY 1260 cacaccacacattaaatgcccgaacacttcggttctgttctggagactgatgctgagtttgt 1319
DB 2226 cacaccacacattaaatgcccgaacacttcggttctgttctggagactgatgctgagtttgt 2285

QY 1320 tagtgact 1379
DB 2286 tagtgact 2345

QY 1380 gctctgcccctctgtgtcacaatcccaacatttatttttcttcttcttcttcttcttcttcttct 1439

Db 2346 g c c t g c c c t g t g t c a c a a t c c c a c e a t t t a t t t t c a t c a a t t t t a c c a t c a a c c a g 2405
 QY 1440 g g c c t g t g c a t c a t a a g t c a c c t c t g a t g c a t c a c a c c a t g g t c t c t c a c c 1499
 Db 2406 g g c c t g t g c a t c a t a a g t c a c c t c t g a t g c a t c a c a c c a t g g t c t c t c a c c 2465
 QY 1500 a t c a c c t t a c c a g t a c c t t a a g a c a t c a t c a t c g c c c a c t t c t g t c a c a a c 1559
 Db 2466 a t c a c c t t a c c a g t a c c t t a a g a c a t c a t c a t c g c c c a c t t c t g t c t c a a a c 2525
 QY 1560 a a c t t g c c a g t c a t c t t c a c c c t g t c c t g c c a g g c a a g g a g a c t c c t g t g a c 1619
 Db 2526 a a c t t g c c a g t c a t c t t c a c c c t g t c c t g c c a g g c a a g g a g a c t c c t g t g a c 2585
 QY 1620 t c a t t t c a g a c a t a a a a t g a c a t a t t c a t g a c a a t g t c c a t a g t a g t c a a a c t g c 1679
 Db 2586 t c a t t t c a g a c a t a a a a t g a c a t a t t c a t g a c a a t g t c c a t a g t a g t c a a a c t g c 2645
 QY 1680 a g a c c t t a a g a c c a g c a c c c a c a c c a a g a g a a g t c a t t a c a g a a t c a a a g t c a c c 1739
 Db 2646 a g a c c t t a a g a c c a g c a c c c a c a c c a a g a g a a g t c a t t a c a g a a t c a a a g t c a c c 2705
 QY 1740 a a g c c t g g t g t c t c t g c c c a g a g t c c a c c a a a g t g t a a c a c a a a c t c t c t t g c c 1799
 Db 2706 a a g c c t g g t g t c t c t g c c c a g a g t c c a c c a a a g t g t a a c a c a a a c t c t c t t g c c 2765
 QY 1800 t c a t c t t a a c a g a g t c t c c a c a g a g a a a c c t t c c a g c a c a a g c a c a a c t t a g c 1859
 Db 2766 t c a t c t t a a c a g a g t c t c c a c a g a g a a a c c t t c c a g c a c a a g c a c a a c t t a g c 2825
 QY 1860 a c a a a t g t c c c a a c t t c a c a c t a c c a t t c t g a a g a c c t c t c a g c c t c t a t a g c a c 1919
 Db 2826 a c a a a t g t c c c a a c t t c a c a c t a c c a t t c t g a a g a c c t c t c a g c c t c t a t a g c a c 2885
 QY 1920 t c c t g c a c c c t g t a a g c a c a g a t c t c t g t c a c t g g c c t a t a g c c g t a c a g a c t a c 1979
 Db 2886 t c c t g c a c c c t g t a a g c a c a g a t c t c t g t c a c t g g c c t a t a g c c g t a c a g a c t a c 2945
 QY 1980 a g c t g a a a a c a g t c t c g t g a c c a t c c t g a a t a c t a g t t c c t c a a t c t c a a c a g a 2039
 Db 2946 a g c t g a a a a c a g t c t c g t g a c c a t c c t g a a t a c t a g t t c c t c a a t c t c a a c a g a 3005
 QY 2040 a g t g g c a c a g c a c a g a a g a a c c g a g t g a t t g t g g a t g t c a c a c t g g a t t g a t c c c 2099
 Db 3006 a g t g g c a c a g c a c a g a a g a a c c g a g t g a t t g t g a t g t a c c a c t g a t t g a t c c c 3065
 QY 2100 t t t g a c c a g t g t a c c a c a t c a g c a a a a g a a t g a c c a a a a g c t t g g c g t t a c a c a g a 2159
 Db 3066 t t t g a c c a g t g t a c c a c a t c a g c a a a a g a a t g a c c a a a g c t t g g c g t t a c a c a g a 3125
 QY 2160 g t a c a g c c a g t t c a g t t c c c t c g a a c a t c c t t c t c c c a a c a c a c a g t t g t t c 2219
 Db 3126 g t a c a g c c a g t t c a g t t c c c t c g a a c a t c c t t c t c c c a a c a c a c a g t t g t t c 3185
 QY 2220 c a c g g t g a a g a c t t g g t c c c a a a t c t g c a c c t t g t g t t c a g a g a g c a c a g t c 2279
 Db 3186 c a c g g t g a a g a c t t g g t c c c a a a t c t g c a c c t t g t g t t c a g a g a g a c a c a g t c 3245
 QY 2280 a c c a a c a a c t g t c t c t t c a g c c t c a g t c a a c a g c t g t g t g t g a a c c c t t g t t c a 2339
 Db 3246 a c c a a c a a c t g t c t c t t c a g c c t c a g t c a a c a g c t g t g t g t g a a c c c t t g t t c a 3305
 QY 2340 c a a t g c g a a t g g t c g c a g a a a c c a g c c g t g g c t a c c a c t g a g t g c c g c c t c 2399
 Db 3306 c a a t g c g a a t g g t c g c a g a a c a c c a g c c g t g g c t a c c a c t g a g t g c c g c c t c 3365
 QY 2400 c t g g c a a g g g a t g a t g c a g t g g a t g a a t g a g t g c t g c a a c c c t g c c a t c 2459
 Db 3366 c t g g c a a g g g a t g a t g c a g t g g a t g a a t g a g t g a g t g c t g c a a c c c t g c c a t c 3425
 QY 2460 c a c a g c a c g t g c a a a t a c t a c a g g a t c c t t t a t c t g c a a a t g c c c g g t g g g t a c c a 2519
 Db 3426 c a c a g c a c g t g c a a c a t a c t a c a g g a t c c t t t a t c t g c a a a t g c c c g g t g g g t a c c a 3485

QY 2520 g t t g g a a a a g g a t a t g c a a t t t g t t a g a a c c t t g t a c a g a g t t t a a a t t a a a g a g 2579
 Db 3486 g t t g g a a a a g g a t a t g c a a t t t g t t a g a a c c t t g t a c a g a g t t t a a a t t a a a g a g 3545
 QY 2580 a a c t t t c t a a t a c a c a c t g t g a a a a c a t t c a g a c c t c a a a a g t t g a a a t g a g a t 2639
 Db 3546 a a c t t t c t a a t a c a c a c t g t g a a a a c a t t c a g a c c t c a a a a g a g t t g a a a t g a g a t 3605
 QY 2640 c a c c a a a c g t t a a a t a t g t g t t t t c a g c g t t a c c t a g t t a c a t c c g a t c a c a g t t c a 2699
 Db 3606 c a c c a a a c g t t a a a t a t g t g t t t t c a g c g t t a c c t a g t t a c a t c c g a t c a c a g t t c a 3665
 QY 2700 c g c t c t a g g a g t c c a a c g c g t g t g a t c t a c t g c a a a a c c t t t c c c t g g c c t c 2759
 Db 3666 c g c t c t a g g a g t c c a a c g c g t g t g a t c t a c t g c a a a c c t t t c c c t g g c c t c 3725
 QY 2760 c a a t g a c c t a t t g a c c t g c t g a t a g a t g c a a a t g t g t c a a c c t c c t g c a a g t c 2819
 Db 3726 c a a t g a c c t a t t g a c c t g c t g a t a g a t g c a a a t g t g t c a a c c t c c t g c a a g t c 3785
 QY 2820 c t c t g t g a g t c t g c a g c t c t t g g a t c t c a g a g c g g a t c t t t a g a g c g g g a c g c t t 2879
 Db 3786 c t c t g t g a g t c t g c a g c t c t t g g a t c t c a g a g c g g a t c t t t a g a g c g g g a c g c t t 3845
 QY 2880 g t g c a a g c g g a a g a g t c c c g a a t g t g a c a a g a c a c c t c c a t c t g c a c t g g a c g g 2939
 Db 3846 g t g c a a g c g g a a g a g t c c c g a a t g t g a c a a g a c a c c t c c a t c t g c a c t g g a c g g 3905
 QY 2940 c g t t g c c c t g c c a g t g c a a g t c g g g a c t c t t c a g t t c a a a a g a t g g a c c a c t c c t g 2999
 Db 3906 c g t t g c c c t g c c a g t g c a a g t c g g g a t a c t t c a g t t c a a c a a g a t g g a c c a c t c c t g 3965
 QY 3000 c c a g a o a t g t g a a g a t g g a t a t a g g c t t g a a a a t g a a a c c t g a a t g a a c c t g a g t t g c c a t t g g 3059
 Db 3966 c c g a g c a t g t g a a g a t g g a t a t a g g c t t g a a a t g a a c c t g a t g a t t g t t a c c t g t t g c c a t t t g g 4025
 QY 3060 c c t t g t g g t c t a a c t g t g a a c c c c t a c a g t t a c a c t g a t c a c t g t g t g a t c g a c g c g c 3119
 Db 4026 c c t t g t g g t c t a a c t g t g a a c c c c t a c a g t t a c a c t g t g t g a t c g a c g c g c 4085
 QY 3120 g g g a g t g g g c t c t c t c a t c c t a g g c a t c g c a c t g a t t g t t a c c t g t t g c a g a a a g a a 3179
 Db 4086 g g g a g t g g g c t c c t g c a t c c t a g g c a t c g c a c t g a t t g t t a c c t g t t g c a g a a a g a a 4145
 QY 3180 t a a a a t g a c a t a a g c a a a c t a t c t t c a a a g t g g a g a t t c c a a a t g t c c c c a t a g c 3239
 Db 4146 t a a a a t g a c a t a a g c a a a c t a t c t t c a a a g t g g a g a t t c c a a a t g t c c c c a t a g c 4205
 QY 3240 t g a t a c c c a a a a t c c t c g t c a c a a g a a t g g g c g a g a a g c t a t t g a a a t g c a t g a 3299
 Db 4206 t g a t a c c c a a a a t c c t c g t c a c a a g a a t t g g g c g a a g c t a t t g a a t g c a t g a 4255
 QY 3300 g a a t g g a a g t a c c a a a a c c t c c c a g a t g a c g a t g t g t a c t a c t c g c c a a g t g t 3359
 Db 4266 g a a t g g a a g t a c c a a a a c c t c c c a g a t g a c g a t g t g t a c t a c t c g c c a a g t g t 4325
 QY 3360 a a g g a a t c c a a a c t t g a a c g a a a c g g a c t a c c c c g c c a c a c t g g a c t g g a c t g c c a g a t c 3419
 Db 4326 a a g g a a t c c a a a c t t g a a c g a a a c g g a c t a c c c g c c a c a c t g g a c t g g a c t g c c a g a t c 4385
 QY 3420 a c g g c a t t c t g a t t t t c c c c g a c a g t a a a c c g t c t t t c a t c a g t a g a a a c a g 3479
 Db 4386 a c g g c a t t c t g a t t t t c c c c g a c a g t a a a c c g t c t t t c a t c a g t a g a a a c a g 4445
 QY 3480 a a g a a g a c t a c t t t t a a g t c c a g a g a g a g g a c t a t t g c t c t g a c c a g t c a c c 3539
 Db 4446 a a g a a g a c t a c t t t t a a g t c c a g a g a g a g a g g a c t a t t g c t c t g a c c a g t c a c c 4505
 QY 3540 t g g g a c c t c t g c a g a g g a c c c a c c a g a g g c t g c g c c a g g a t t t g t c g g a g c c a c 3599
 Db 4506 t g g g a c c t c t g c a g a g g a c c c a c c a g a g g c t g c g c c a g g a t t t g t c g g a g c c a c 4565

Qy	3600	gctgagtggcaagcaggagaagggacaggcatgcggyggcgtagaccacagtggaggagaca	3650
Db	4566	gctgagtggcaagcaggagaagggacaggcatgcggyggcgtagaccacagtggaggagaca	4625
Qy	3660	ggttgatgtggaaaccacaggctgctcaattcacgaacctttgttgttactgtgaacgtgaat	3719
Db	4626	ggttgatgtggaaaccacaggctgctcaattcacgaacctttgttgttactgtgaacgtgaat	4685
Qy	3720	gtggggccagtatacaagagagtcctctgagtgactgcaccatggcactggcacagagcg	3779
Db	4686	gtggggccagtatacaagagagtcctctgagtgactgcaccatggcactggcacagagcg	4745
Qy	3780	actattagccagggcagacacactagacttcagtcgagggacctggtttccccttcgtttg	3839
Db	4746	actattagccagggcagacacactagacttcagtcgagggacctggtttccccttcgtttg	4805
Qy	3840	cacttttagtaaatgtgggtgggagggtttcccttttgatctgttttgagactgttccagaaa	3899
Db	4806	cacttttagtaaatgtgggtgggagggtttcccttttgatctgttttgagactgttccagaaa	4865
Qy	3900	gaaggtctcccttccccgaacacttcocatagcagcaaatgttgatgccattttgatccacaa	3959
Db	4866	gaaggtctcccttccccgaacacttcocatagcagcaaatgttgatgccattttgatccacaa	4925
Qy	3960	aatactgctgtttaattatttccctgccagcgctgcgtgcataaacacagatgagga	4019
Db	4926	aatactgctgtttaattatttccctgccagcgctgcgtgcataaacacagatgagga	4985
Qy	4020	tgagcgtaccaactgaagtcotgaagatgtgcgccattgaacggacagtgttttcatactgtt	4079
Db	4986	tgagcgtaccaactgaagtcotgaagatgtgcgccattgaacggacagtgttttcatactgtt	5045
Qy	4080	ctaagttgtcttatgtctacagtttccaagcagccccacagtcagggaaaatgtgtgaggc	4139
Db	5046	ctaagttgtcttatgtctacagtttccaagcagccccacagtcagggaaaatgtgtgaggc	5105
Qy	4140	accgcacacaactgcgaatgtgttttttaagtcaaagtgacacatgtatttaagattttt	4199
Db	5106	accgcacacaactgcgaatgtgttttttaagtcaaagtgacacatgtatttaagattttt	5165
Qy	4200	tttaaatctcttttgcagttaaactcactttttccaacaagccttgatacagggcaaac	4259
Db	5166	tttaaatctctytgcagttaaactcacttttccaacaagccttgatacagggcaaac	5225
Qy	4260	aacttatattgttttagctgagagctcagcagcacagatitgcaggcagggggcacctt	4319
Db	5226	aacttatattgttttagctgagagctcagcagcacagatitgcaggcagggggcacctt	5285
Qy	4320	tcatactagagggcccagcctgggacctggagctctgatcaccaattgltggagggccagag	4379
Db	5286	tcatactagagggcccagcctgggacctggagctctgatcaccaattgltggagggccagag	5345
Qy	4380	cagctgcgatgagagagaaaaatgtaaacctgaacgcaggtttcacccactctagaaaagca	4439
Db	5346	camctgcgatgagagagaaaaatgtaaacctgaacgcaggtttcacccactctagaaaagca	5405
Qy	4440	gcttgctgagccccctcaactgatatgtgttagagggatggctgaaatagccaggttcaga	4499
Db	5406	gcttgctgagccccctcaactgatatgtgttagagggatggctgaaatagccaggttcaga	5465
Qy	4500	tttctcgtatacaacagtgtgttgggaagctgtgtggaattcctgagggagaacacaggagcc	4559
Db	5466	tttctcgtatacaacagtgtgttgggaasctgtgtggaattcctgagggagaacacaggagcc	5525
Qy	4560	gagatgagccacacatgagtttgcataccgggctactgcagocactttgtatccccgaatct	4619
Db	5526	gagatgagccacacatgagtttgcataccgggctactgcagocactttgtatccccgaatct	5585
Qy	4620	catgttcacaaaaacccatgtaaaccttcaaacactcaagctgtttattcggctgaaaga	4679
Db	5586	catgttcacaaaaacccatgtaaaccttcaaacactcaagctgtttattcggctgaaaga	5645
Qy	4680	ataacttttttttctcacccagtcattttgacctcttcataatggctgtgtgcacacctcc	4739

[illegible]

PD 15-DEC-1998.
 XX
 PF 15-MAR-1996; 96US-0616844.
 XX
 PR 09-FEB-1996; 96US-0599654.
 PR 10-FEB-1995; 95US-0386844.
 PR 07-JUN-1995; 95US-0458873.
 PR 15-MAR-1996; 96US-0616844.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Falb DA;
 XX
 DR WPI: 1999-069743/06.
 DR P-PSDB; AAW89299.
 XX
 PT DNA encoding rchd528 polypeptide - associated with cardiovascular
 PT disease
 XX
 XX Claim 2; Fig 30; 122pp; English.
 XX
 CC The present sequence encodes rchd528 protein. A method has been
 CC developed for producing the rchd528 gene product. The present invention
 CC describes methods and compositions for the treatment and diagnosis of
 CC cardiovascular diseases, including: atherosclerosis; ischaemia;
 CC restenosis; reperfusion; hypertension; and arterial inflammation.
 XX
 SQ Sequence 6407 BP; 1594 A; 1762 C; 1520 G; 1509 T; 22 other;

Query Match 73.3%; Score 5098.4; DB 20; Length 6407;
 Best Local Similarity 94.3%; Pred. No. 0;
 Matches 5412; Conservative 21; Mismatches 8; Indels 301; Gaps 2;

QY 1 gttcgtgaaagaattccgcgttttcaaaacaaagatggaacagcctcggagatggaac 60
 DB 666 gttcgtgaaagaattccgcgttttcaaaacaaagatggaacagcctcggagatggaac 725
 QY 61 agagagggcgtggtgtcagaagaatggaactgtgcacagccaaagagccaccattc 120
 DB 726 agagagggcgtggtgtcagaagaatggaactgtgcacagccaaagagccaccattc 785
 QY 121 ggcttgagccttcccttctcgtcttggagatggagagctgaccagccttcttag 180
 DB 786 ggcttgagccttcccttctcgtcttggagatggagagctgaccagccttcttag 845
 QY 181 gaagagaaattcctcaggaaccagatctctcctggctgcatttctacaggaacagcttc 240
 DB 846 gaagagaaattcctcaggaaccagatctctcctggctgcatttctacaggaacagcttc 905
 QY 241 ctctcctcttagaccttctcctcactcttgaaagtacagagaagcttaacaactcac 300
 DB 906 ctctcctcttagaccttctcctcactcttgaaagtacagagaagcttaacaactcac 965
 QY 301 tggcctccagagctcctcagtcagtcacaaacaaagacatgttctacccgtgttcac 360
 DB 966 tggcctccagagctcctcagtcagtcacaaacaaagacatgttctacccgtgttcac 1025
 QY 361 tga tgg tggcccgagaacgtcgatcttggaggtcagtcctggacctgtgagcaagac 420
 DB 1026 tga tgg tggcccgagaacgtcgatcttggaggtcagtcctggacctgtgagcaagac 1085
 QY 421 agaaggttcccaagacctccagaattgccaagacttcaatcctcagtccttcttccac 480
 DB 1086 agaaggttcccaagacctccagaattgccaagacttcaatcctcagtccttcttccac 1145
 QY 481 ctctcagtggaatcgagagaacacagtagaagtaactggaaatccagggtatggaatt 540
 DB 1146 ctctcagtggaatcgagagaacacagtagaagtaactggaaatccagggtatggaatt 1205
 QY 541 cattgaaccatccacagaaaatgaatttgacttacgtcttcttgcg-tggcaaaatgattc 599
 DB 1206 cattgaaccatccacagaaaatgaatttgacttacgtcttcttgcg-tggcaaaatgattc 1265

QY 600 ccaaacctttggagaacatcagcttgccagcagctctgaggtgcaaaatggaagtcccat 659
 DB 1266 ccaaacctttggagaacatcagcttgccagcagctctgaggtgcaaaatggaagtcccat 1325
 QY 660 gtctcagactgagactgtgtctaggtcagtcgaccccatgagaggtgagagatacactgc 719
 DB 1326 gtctcagactgagactgtgtctaggtcagtcgaccccatgagaggtgagagatacactgc 1385
 QY 720 aactggctcttgaccacacagacacacatctgcagatgtgcaggaagctctgttccata 779
 DB 1386 aactggctcttgaccacacagacacacatctgcagatgtgcaggaagctctgttccata 1445
 QY 780 tcttgaagggtgtaagtcttcaagtgttgacccagttctcagactctactgtacagtcgg 839
 DB 1446 tcttgaagggtgtaagtcttcaagtgttgacccagttctcagactctactgtacagtcgg 1505
 QY 840 aggaagtcacacagcattgggagataggagttattcagagttcttcatcatcttctctc 899
 DB 1506 aggaagtcacacagcattgggagataggagttattcagagttcttcatcatcttctctc 1565
 QY 900 ggaagcttgaattcattcagcaccacgtggagaaacttcaa----- 940
 DB 1566 ggaagcttgaattcattcagcaccacgtggagaaacttcaaacttggagagacgcgaga 1625
 QY 941 ----- 940
 DB 1626 gccaggccaagcactaggtgacagttccgcgaatgcagagagacaggaacttctgggtgccc 1685
 QY 941 ----- 940
 DB 1686 ctctctcgccaccacacacttggctactgtcacttgaaacggggaacgcacacactgcggtc 1745
 QY 941 ----- 940
 DB 1746 tgtcacctcaccacacacacagcatgacagacttctggggaagcagcagccctgcagc 1805
 QY 941 ----- 940
 DB 1806 ggccatgccccaaagaacagaggggtgctctctctgcagtaaacgtgacggagcagatggg 1865
 QY 941 -----tcgtgggattagctacgg 959
 DB 1866 cctggtctcacggtcactggcgcctccagtcgactcggagtcgctgggattagctacgg 1925
 QY 960 tcaagtcgctggcacagctattgaaacaaagacttccagagaccacacagaccacaccta 1019
 DB 1926 tcaagtcgctggcacagctattgaaacaaagacttccagagaccacacagaccacaccta 1985
 QY 1020 cctgtcacttacttccacaaagagaaacgggcttacttccattacagataacagttc 1079
 DB 1986 cctgtcacttacttccacaaagagaaacgggcttacttccattacagataacagttc 2045
 QY 1080 atctcagacattgtggagagctcaacttcttattataaaatctcaaaacttccacattc 1139
 DB 2046 atctcagacattgtggagagctcaacttcttattataaaatctcaaaacttccacattc 2105
 QY 1140 agagtaatcctccttctcctcagtcagactgagagaagtaactctcctcctatgcaggg 1199
 DB 2106 agagtaatcctccttctcctcagtcagactgagagaagtaactctcctcctatgcaggg 2165
 QY 1200 ggaataatgctcagccttctactgagtcgccaagtctgcatacatccaacttccgctcta 1259
 DB 2166 ggaataatgctcagccttctactgagtcgccaagtctgcatacatccaacttccgctcta 2225
 QY 1260 cacaccacacattaatgcccgaacacttcggttgttcttgacactgatgctgagttgt 1319
 DB 2226 cacaccacacattaatgcccgaacacttcggttgttcttgacactgatgctgagttgt 2285
 QY 1320 tagtgactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctt 1379
 DB 2286 tagtgactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctt 2345

Db 4506 tgggacctctgctcagaggaccgcaaccaggaggctgcgccaggatttgtcgggagccac 4565
Qy 3600 gctgagtggaagcagagaagaggacaggcatcgggggcgtgaccacagtggaagagaca 3659
Db 4566 gctgagtggaagcagagaagaggacaggcatcgggggcgtgaccacagtggaagagaca 4625
Qy 3660 ggtgagtggaagcagagcgtgctcattcagacaccttgtgttactgtgaacgtgaat 3719
Db 4626 ggtgagtggaagcagagcgtgctcattcagacaccttgtgttactgtgaacgtgaat 4685
Qy 3720 gtgggcccagtatcaagagagttctctctgtagtgcactgcaacctggcactggcaccaggcg 3779
Db 4686 gtgggcccagtatcaagagagttctctctgtagtgcactgcaacctggcactggcaccaggcg 4745
Qy 3780 actattagccagggcagaccactagacttcagtgacagggaccctgtttcccttcggttg 3839
Db 4746 actattagccagggcagaccactagacttcagtgacagggaccctgtttcccttcggttg 4805
Qy 3840 cactttagtaaatgggtggagggttcccttttgatctgttttgagactgtccagaaa 3899
Db 4806 cactttagtaaatgggtggagggttcccttttgatctgttttgagactgttccagaaa 4865
Qy 3900 gaagggttcctttcccagagacacttccataggcagcaaatttggtgatttcatttcagcaa 3959
Db 4866 gaagggttcctttcccagagacacttccataggcagcaaatttggtgatttcatttcagcaa 4925
Qy 3960 aatactggtgttaattatttctcgtccagggcctcgtgctataaacacagatgagga 4019
Db 4926 aatactggtgttaattatttctcgtccagggcctcgtgctataaacacagatgagga 4985
Qy 4020 tgagcgtaccactgaagtgatgcaccttgacgagcagtgatttttcatatgttt 4079
Db 4986 tgascgtaccactgaagtcgaagatgcaccttgacgagcagtgatttttcatatgttt 5045
Qy 4080 ctagggtgttattgctacagtttccaaagccagcccccacagtgaggaaatgtgtgagcc 4139
Db 5046 ctagggtgttattgctacagtttccaaagccacccccacagtgaggaaatgtgtgagcc 5105
Qy 4140 accgcacacaaactgcaatgtgtttttaagtcgaagtgacacatgtatttaagaattttt 4199
Db 5106 accgcacacaaactgcaatgtgtttttaagtcgaagtgacacatgtatttaagaattttt 5165
Qy 4200 tttaaaatctcttgcaagtaaatctcaacttttcaaacagcctgcatcagggcacaaac 4259
Db 5166 tttaaaatctcttgcaagtaaatctcaacttttcaaacagcctgcatcagggcacaaac 5225
Qy 4260 aacttatatttggttttagctgaggtcagcagggcagattgagcaggggggacattt 4319
Db 5226 aacttatatttggttttagctgaggtcagcagggcagattgagcaggggggacattt 5285
Qy 4320 tcatccatagggcccagcctgggctggactcgtgactcgtatcaccaattgtgagggccaggg 4379
Db 5286 tcatccatagggcccagcctgggctggactcgtgactcgtatcaccaattgtgagggccaggg 5345
Qy 4380 cagctgctatggaggagaaatgtcaaacgtgaacgaggtttccaccacttaggaaagca 4439
Db 5346 camctgctatggaggagaaatgtcaaacgtgaacgaggtttccaccacttaggaaagca 5405
Qy 4440 gcttgttagccccctgcagctgcatgtggttagagggaatgggtgaatagcaggttaga 4499
Db 5406 gcttgttagccccctgcasctgcatgtggttagagggaatgggtgaatagcaggttaga 5465
Qy 4500 ttctctgcatcaacagtgctttggggaagctgtgtgattcctgaggaagaacaggagcc 4559
Db 5466 ttctctgcatcaacagtgctttggggaagctgtgtgattcctgaggaagaacaggagcc 5525
Qy 4560 gagatggagccacacatgagtttgctacccgctactgcagcactttgtaccaccagactct 4619
Db 5526 gagatggagccacacatgaaatggtccaccgctactgcagcactttgtaccaccagactct 5585
Qy 4620 catgtccaaaaaccccatgttaaactttcaaccactcaaaagctgtttattcggtgaagaa 4679
Db 5586 catgtccaaaaaccccatgttaaactttcaaccactcaaaagctgtttattcggtgaagaa 5645

Qy 4680 ataacttttttttttccaccacagtcatttgttacctcttccatagctgtgtcgcaccctcc 4739
Db 5646 ataacttttttttccaccacagtcatttgttacctcttccatagctgtgtcgcaccctcc 5705
Qy 4740 agaaacgtgggttacttccagtcagtggtggagaactgaagactccggttgcgaggg 4799
Db 5706 agaaacgtgggttacttccagtcagtggtggagaactgaagactccggttgcgaggg 5765
Qy 4800 aactgaggggttgaccttogggaagagtgctccactcatcttatttattatccctgtgatg 4859
Db 5766 aactgaggggttgaccttogggaagagtgctccactcatcttatttattatccctgtgatg 5825
Qy 4860 tgggtcctgcaggagagacatccagtaactcgtgtctttaaattggcaccctggggaactgt 4919
Db 5826 tgggtcctgcaggagagacatccagtaactcgtgtctttaaattggcaccctggggaactgt 5885
Qy 4920 gtttattggccttcttctggggcatcctcgttttgatgaagtgaaggaacacagagta 4979
Db 5886 gtttattggccttcttctggggcatcctcgtktcggatgaagtgaaggaacacagagta 5945
Qy 4980 aaagaattgtctccacctgaagcggggagtcctcgccttcacatttctggaatggtgcag 5039
Db 5946 aaagaattgtctccacctgaagcggggagtcctcgccttcacatttctggaatggtgcag 6005
Qy 5040 ccaactggggacagttctgccccgggcatggtgttttccaaggtcctcctataataatc 5099
Db 6006 ccaactggggacagttctgccccgggcatggtgttttccaaggtcctcctataataatc 6065
Qy 5100 cctattctacataatccttggccctgatggttttaagcaagaactcctctgtccccatgg 5159
Db 6066 cctattctacataatccttggccctgatggttttaagcaagaactcctctgtccccatgg 6125
Qy 5160 tctccacctccaccatccacctcgtctgtagcaagtccttagtcaggggaggtgcatttt 5219
Db 6126 tctccacctccaccatccacctcgtctgtagcaagtccttagtcaggggaggtgcatttt 6185
Qy 5220 agtagttaaatggacattatccatgagataaaataaaggagaaactgttttatcagtgga 5279
Db 6186 agtagttaaatggacattatccatgagataaaataaaggagaaactgttttatcagtgga 6245
Qy 5280 ggctaaactaaaatttcaagtgctgccttttgaatacttggcctctctctctctctaga 5339
Db 6246 ggctaaactaaaatttcaagtgctgccttttgaatacttggcctctctctctctctaga 6305
Qy 5340 accaatggcccttgtgtgctacggcctcgcacctaactgagaggttctgagctcctgca 5399
Db 6306 accaatggcccttgtgtgctacggcctcgcacctaactgagaggttctgagctcctgca 6365
Qy 5400 gctcactgagcccaagactaggcttcttggtccttccgc 5441
Db 6366 gctcactgagcccaagactaggcttcttggtccttccgc 6407

RESULT 3

AA88578

ID AAA88578 standard; cDNA; 6407 BP.

XX

AC AAA88578;

XX

DT 05-FEB-2001 (first entry)

XX

DE Human rchd528 cDNA.

XX

KW Human; rchd528 gene; differential expression; HUVEC; shear stress;
endothelial cell; cardiovascular disease; inflammation;
atherosclerosis; antiinflammatory; antiatherosclerotic; diagnosis;
gene therapy; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 19..4464

Db 4146 taaaaatgacataagcaaaactatcttccaaaagtggagatttccaaaatgtccccatatgc 4205
Qy 3240 tgaatacccacaaaaatcctcgctcacagaatgggcccgaagaagctattgaatatgcata 3299
Db 4206 tgaatacccacaaaaatcctcgctcacagaatgggcccgaagaagctattgaatatgcata 4205
Qy 3300 gaa tggaaagtaccacaaaaactcctccacagatgacggatgtgtactactcgcctacaaagt 3359
Db 4266 gaa tggaaagtaccacaaaaactcctccacagatgacggatgtgtactactcgcctacaaagt 4325
Qy 3360 aaggaatccagaaacttgaagaaacggactactacccggcctacactggactgcccagatc 3419
Db 4326 aaggaatccagaaacttgaagaaacggactactacccggcctacactggactgcccagatc 4385
Qy 3420 acggcatcttgcatttcccccagacagtataaacccgtcttctcatcagtgatgaaagcag 3479
Db 4386 acggcatcttgcatttcccccagacagtataaacccgtcttctcatcagtgatgaaagcag 4445
Qy 3480 aagaagagactacttttaagtccaggagagagaggactcattgtctgtgagccagtacc 3539
Db 4446 aagaagagactacttttaagtccaggagagagaggactcattgtctgtgagccagtacc 4505
Qy 3540 tgggaactctgtctcagaggaccgacacaggaggtgcgccagagatttgcggagagccac 3599
Db 4506 tgggaactctgtctcagaggaccgacacaggaggtgcgccagagatttgcggagagccac 4565
Qy 3600 gctgagtggcaagcaggaagagagacaggcatgcggggcgtgacacacagtggaagagaca 3659
Db 4566 gctgagtggcaagcaggaagagagacaggcatgcggggcgtgacacacagtggaagagaca 4625
Qy 3660 ggtggatggtggaacccacaggctgcctcatcagcaaccttgtgttactgtgaacgtgaat 3719
Db 4626 ggtggatggtggaacccacaggctgcctcatcagcaaccttgtgttactgtgaacgtgaat 4685
Qy 3720 gtgggcagatcaagagagtctctgtgagtactgcacacatggcactggcaccaggcgg 3779
Db 4686 gtgggcagatcaagagagtctctgtgagtactgcacacatggcactggcaccaggcgg 4745
Qy 3780 actattagccagggcagaccactagactcagtgccaggagcctggtttcccttcgtttg 3839
Db 4746 actattagccagggcagaccactagactcagtgccaggagcctggtttcccttcgtttg 4805
Qy 3840 cacttttagtaaaatgggtgggaggtttcccttttggatctgttttgagactggttccagaaa 3899
Db 4806 cacttttagtaaaatgggtgggaggtttcccttttggatctgttttgagactggttccagaaa 4865
Qy 3900 gaaggcttccttcccgagacacttccataggaagcaatttgggtgatctatttgcagcaa 3959
Db 4866 gaaggcttccttcccgagacacttccataggaagcaatttgggtgatctatttgcagcaa 4925
Qy 3960 aactactggctgttaattattttctctgccagcgctgcgtgcttaaacacacagatgagga 4019
Db 4926 aactactggctgttaattattttctctgccagcgctgcgtgcttaaacacacagatgagga 4985
Qy 4020 tgaagctaccactgaagctgtaagatgtcgccatttgaacggacagtggttttccatgatgtt 4079
Db 4986 tgaagctaccactgaagctgtaagatgtcgccatttgaacggacagtggttttccatgatgtt 5045
Qy 4080 ctagggtgttattgtctacagtttccaaagccagcccccacacagtgagaaatgtgtgaggc 4139
Db 5046 ctagggtgttattgtctacagtttccaaagccagcccccacacagtgagaaatgtgtgaggc 5105
Qy 4140 accgcacacactgcaatgtgttttttaagtcaaggtgacacatgtatttaagatttttt 4199
Db 5106 accgcacacactgcaatgtgttttttaagtcaaggtgacacatgtatttaagatttttt 5165
Qy 4200 tttaaaaatctcttgcagtttaaaatctcacctttttcaaacaaagcctggatcaggggcaaac 4259
Db 5166 tttaaaaatctcttgcagtttaaaatctcacctttttcaaacaaagcctggatcaggggcaaac 5225
Qy 4260 aacttatatttggttttagctggaggctcagcagagcagatgtgacggcagggggacattt 4319
Db 5226 aacttatatttggttttagctggaggctcagcagagcagatgtgacggcagggggacattt 5285

Qy 4320 tcaccatgagggcccagcctggggcctggactctgatcaccattgtggaggccagagg 4379
Db 5286 tcaccatgagggcccagcctggggcctggactctgatcaccattgtggaggccagagg 5345
Qy 4380 cagctgcgta tgggaggaataatgtcaaaactgaacgcagggtttaccactctagaaaaa 4439
Db 5346 camctgcgta tgggaggaataatgtcaaaactgaacgcagggtttaccactctagaaaaa 5405
Qy 4440 gcttgttgagcccctgcagctggatgtggttagagggtaggctgaaatagcaggttaga 4499
Db 5406 gcttgttgagcccctgcascctggatgtggttagagggtaggctgaaatagcaggttaga 5465
Qy 4500 tttcctgcatacaacagtgcttttgggaagctgttggattcctgaggaagacacaggagcc 4559
Db 5466 tttcctgcatacaacagtgcttttgggaasactgttggattcctgaggaagaacaggagcc 5525
Qy 4560 gagatggagccacacatgagtttgcctcacccggctactgcagcaactttgtaccacaaatct 4619
Db 5526 gagatggagccacacatgaaattgtctcacggctactgcagcaactttgtaccacaaatct 5585
Qy 4620 catctcacaaaaccccatgtaaaactttcaaccactcaaaagctgtttattcggctgaa 4679
Db 5586 catgtcacaaaaccccatgtaaaactttcaaccactcaaaagctgtttattcggctgaa 5645
Qy 4680 ataacttttttttcaaccagtcatttgtacctcttcataatggctgtgcgacccctcc 4739
Db 5646 ataacttttttttcaaccagtcatttgtacctcttcataatggctgtgcgacccctcc 5705
Qy 4740 agaaactgggttatacttccagtcagtggtgggaactgaagacttcoggttggctgagg 4799
Db 5706 agaaactgggttatacttccagtcagtggtgggaactgaagacttcoggttggctgagg 5765
Qy 4800 aactgaggggttgaccttcgggaagagttccactcatcttatttattatcctgtgatg 4859
Db 5766 aactgaggggttgaccttcgggaagagttccactcatcttatttattatcctgtgatg 5825
Qy 4860 tgggtcctgcaggagagacatccagtcactcgtgtctttaaattgccaactgggaactcgt 4919
Db 5826 tgggtcctgcaggagagacatccagtcactcgtgtctttaaattgccaactgggaactcgt 5885
Qy 4920 gtttattggcctcttttggggcatactcctgggttttggatgaagtggagggaatcacagagata 4979
Db 5886 gtttattggcctcttttggggcatactcctgggtkttcgatgaagtggagggaatcacagagata 5945
Qy 4980 aagaataatgtctccacactgaagcggggagtcgccgttccacttcttggaatggtgcag 5039
Db 5946 aagaataatgtctccacactgaagcggggagtcgccgttccacttcttggaatggtgcag 6005
Qy 5040 ccactggggacagttctgcgccgggcatggtgttttctcaagggtccctctaaataataatc 5099
Db 6006 ccactggggacagttctgcgccgggcatggtgttttctcaagggtccctctaaataataatc 6065
Qy 5100 cctatttttaacataacttgcgcctgatgggtttttaagcaagaactcctgtgctccatg 5159
Db 6066 cctatttttaacataacttgcgcctgatgggtttttaagcaagaactcctgtgctccatg 6125
Qy 5160 tctccacactcacacacccctcgttagcaagtccttagcaggtcagggaggtgcatttt 5219
Db 6126 tctccacactcacacacccctcgttagcaagtccttagcaggtcagggaggtgcatttt 6185
Qy 5220 agtsgtttaaaattgcacttatccatgagataataaaaaagagaactgtttttatcagtgga 5279
Db 6186 agtsgtttaaaattgcacttatccatgagataataaaaaagagaactgtttttatcagtgga 6245
Qy 5280 ggcataactaaaaattcaaaagtgcgcctttttgaaatcttgggcctctctctctctgtaga 5339
Db 6246 ggcataactaaaaattcaaaagtgcgcctttttgaaatcttgggcctctctctctctgtaga 6305
Qy 5340 accaaatggcctttgttggtcagcggcctgcacactaactggagaggttctgagctcctgca 5399
Db 6306 accaaatggcctttgttggtcagcggcctgcacactaactggagaggttctgagctcctgca 6365

QY 5400 gctcacctgagccacagactaggtcttctggtcctctccgc 5441
|||||
Db 6366 gctcacctgagccacagactaggtcttctggtcctctccgc 6407

RESULT 4

AA289796
ID AA289796 standard; cDNA; 6407 BP.

AC AA289796;

XX 05-MAY-2000 (first entry)

XX Human cardiovascular disease associated gene rchd528.

XX Differentially expressed; cardiovascular disease; atherosclerosis;
KW ischaemia; reperfusion; hypertension; restenosis; arterial inflammation;
KW rchd528; transmembrane protein; ss.

XX Homo sapiens.

XX US6020463-A.

XX 01-FEB-2000.

XX 06-OCT-1997; 97US-0944423.

XX 09-FEB-1996; 96US-0599654.

XX 10-FEB-1995; 95US-0386844.

XX 07-JUN-1995; 95US-0485573.

XX (BGHM) BRIGHAM & WOMENS HOSPITAL.

PA (MILL-) MILLENNIUM PHARM INC.

XX Glanbrone MA, Falb DA;

XX WPI; 2000-146911/13.

DR P-PSDB; AAY78508.

XX Marker proteins for the diagnosis of cardiovascular diseases such as

PT atherosclerosis and hypertension, comprising peptide sequences derived

PT from the rchd523 transmembrane protein -

XX Examples; Fig 30; 121pp; English.

XX This sequence represents the human rchd528 gene sequence. This sequence
CC is related to the rchd523 transmembrane polypeptide encoded by cDNA
CC contained in the plasmid pfchd523. The rchd523 protein is differentially
CC expressed in diseased cells compared to healthy cells. The rchd523
CC protein may be used as a marker protein for the diagnosis of
CC cardiovascular diseases including atherosclerosis, ischaemia,
CC reperfusion, hypertension, restenosis and arterial inflammation. rchd523
CC peptides may be used as antigens in the production of antibodies specific
CC for rchd523. The anti-rchd523 antibodies may then be used in diagnostic
CC assays to quantitate rchd523 peptides in samples.

XX Sequence 6407 BP; 1594 A; 1762 C; 1520 G; 1509 T; 22 other;

Query Match

Best Local Similarity 94.3%; Score 5098.4; DB 21; Length 6407;
Matches 5412; Conservative 21; Mismatches 8; Indels 301; Gaps 2;

QY 1 gttcgtgaaagaattgccgttttcaacaaagagtggaacgcctcggagatgggaac 60
|||||
Db 666 gttcgtgaaagaattgccgttttcaacaaagagtggaacgcctcggagatgggaac 725
|||||
QY 61 agagagggcagatgggctgtcagaagaatgactgtgcacagccaagggccacattc 120
|||||
Db 726 agagagggcagatgggctgtcagaagaatgactgtgcacagccaagggccacattc 785
|||||
QY 121 ggcttgagccctctcttctctctgttggagatggagagctgaccagcctcttag 180
|||||

Db 786 ggcttgagcccgctcttcttctctggttgagagatggagagctgacacagccttctag 845
QY 181 gaagagaaattctctcagagaccagatctctctcgtgctatttctacagagacagcttc 240
|||||
Db 846 gaagagaaattctctcagagaccagatctctctcgtgctatttctacagagacagcttc 905
|||||
QY 241 ctctcctcttagacatttctcacccttgaagagtcagagagcttaacaactccac 300
|||||
Db 906 ctctcctcttagacatttctcacccttgaagagtcagagagcttaacaactccac 965
|||||
QY 301 tggctccagagctcctcagtcagtcacaacaaagacaatgcattgttgcaccgtgttcac 360
|||||
Db 966 tggctccagagctcctcagtcagtcacaacaaagacaatgcattgttgcaccgtgttcac 1025
|||||
QY 361 tgatggtggccgagagacgctgcgtatctttgaagtcagtcagtcggacacctgtgagaagac 420
|||||
Db 1026 tgatggtggccgagagacgctgcgtatctttgacggtcagtcgtggacacctgtgagaagac 1085
|||||
QY 421 agaaggtctcccaagagactccagaattgcacagacttcacctcagtcctctcttcacc 480
|||||
Db 1086 agaaggtctcccaagagactccagaattgcacagacttcacctcagtcctctcttcacc 1145
|||||
QY 481 ctctgcagtggaaatcgagaagaacagtagagtaactgggaatccaggggatgagaatt 540
|||||
Db 1146 ctctgcagtggaaatcgagaagaacagtagagtaactgggaatccaggggatgagaatt 1205
|||||
QY 541 cattgaaccatccacagagaaaatgaatttggacttaactgttcttgcg-tggcaaaatgattc 599
|||||
Db 1206 cattgaaccatccacagagaaaatgaatttggacttaactgttcttgcg-tggcaaaatgattc 1265
|||||
QY 600 cccaaccttggagaacatcagcttgcagcagctctgagtgcaaaatcggaagtcacct 659
|||||
Db 1266 cccaaccttggagaacatcagcttgcagcagctctgagtgcaaaatcggaagtcacct 1325
|||||
QY 660 gtctcagactgagactgtgtctagtcagtcgacccatcagagtgagtgagagatcactgc 719
|||||
Db 1326 gtctcagactgagactgtgtctagtcagtcgacccatcagagtgagtgagagatcactgc 1385
|||||
QY 720 aactggcttcttgaccaacagacacaaatctgcagatgtgacaggaagctctgcttcata 779
Db 1386 aactggcttcttgaccaacagacacaaatctgcagatgtgacaggaagctctgcttcata 1445
|||||
QY 780 tctgaaagtgatgaatgcttcagttgacccagttctcagactctactgtacagctcgg 839
|||||
Db 1446 tctgaaagtgatgaatgcttcagttgacccagttctcagactctactgtacagctcgg 1505
|||||
QY 840 aggaagtcacacagatctggagatagagttattcagagttctcattcatctctctc 899
Db 1506 aggaagtcacacagatctggagatagagttattcagagttctcattcatctctctc 1565
|||||
QY 900 ggaagcttgaattcattcagcaccacagctggagaacgttcaa----- 940
|||||
Db 1566 ggaagcttgaattcattcagcaccacagctggagaacgttcaaaccttggaaagacagccgaga 1625
QY 941 ----- 940
Db 1626 gccagggccaaagcactagtgacgttcgcgccaatgcagagagacagacttctggggtgcc 1685
QY 941 ----- 940
Db 1686 ctctctcggcaccacacaccttggctactgtcactggaaacgggggaacgcacacactcggtc 1745
QY 941 ----- 940
Db 1746 tgtcacctccacaacacacagcatgacacgacttctgtgggaagcaggcagccctgcagc 1805
QY 941 ----- 940
Db 1806 ggccatgcccccaagaacacagaggtgtcctctctgcacgtaaacgtgacggagacatggg 1865
QY 941 -----tcgctgggattagctacgg 959
|||||
Db 1866 cctggtctcaggtcacttgcgcgctccagtgactgcagtcgctgggattagctacgg 1925

Qy 960 tcaagtgcgtgagcagctattgaacaaaggacttcagcgaccacacagacacaccta 1019
 Db 1926 tcaagtgcgtgagcagctattgaacaaaggacttcagcgaccacacagacacaccta 1985
 Qy 1020 cctgtcatctacttccaccaaggagaaacggcgcttactgtccattacagataacagttc 1079
 Db 1986 cctgtcatctacttccaccaaggagaaacggcgcttactgtccattacagataacagttc 2045
 Qy 1080 atcctcagacaattgtgagagactcaaatcttatattataaactccaactcttcacattc 1139
 Db 2046 atcctcagacaattgtgagagactcaaatcttatattataaactccaactcttcacattc 2105
 Qy 1140 agagtattcctccctttctcatgctcagactgagagaaaglaaacatctcatcctatgacgg 1199
 Db 2106 agagtattcctccctttctcatgctcagactgagagaaaglaaacatctcatcctatgacgg 2165
 Qy 1200 ggaatatgctcagcctctctactgagtcgagcttctgcatalacatcccaaccttcgctccta 1259
 Db 2166 ggaatatgctcagcctctctactgagtcgagcttctgcatalacatcccaaccttcgctccta 2225
 Qy 1260 cacaccacaactaatatgcgaacacttcggttctgtctggaactgatgctgagtttgt 1319
 Db 2226 cacaccacaactaatatgcgaacacttcggttctgtctggaactgatgctgagtttgt 2285
 Qy 1320 tagtgactcctcctcctctctcctcctcctcctcctcctcctcctcctcctcctcctt 1379
 Db 2286 tagtgactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctt 2345
 Qy 1380 gccctcgcctctgttcaacaatcccaacctttatttttcatcaattttacccaacaacag 1439
 Db 2346 gccctcgcctctgttcaacaatcccaacctttatttttcatcaattttacccaacaacag 2405
 Qy 1440 ggcctctgtcatctactaaagtctactcctgatgcatcacaacacttggtctcctcacc 1499
 Db 2406 ggcctctgtcatctactaaagtctactcctga tgcatacacaacttggtctcctcacc 2465
 Qy 1500 atcacctttaccagtaacctaaagacatactacatctgcccactttctgtctcacaacac 1559
 Db 2466 atcacctttaccagtaacctaaagacatactacatctgcccactttctgtctcacaacac 2525
 Qy 1560 aaccttgccacagtcaattcttaccctcgttccctgcccagggaaggagactcctgtgac 1619
 Db 2526 aaccttgccacagtcaattcttaccctcgttccctgcccagggaaggagactcctgtgac 2585
 Qy 1620 ttcatttcagacatcaacaatgacatcatctcatgacaatgctccatagtagtcaaatgc 1679
 Db 2586 ttcatttcagacatcaacaatgacatcatctcatgacaatgctccatagtagtcaaatgc 2645
 Qy 1680 agaccttaagagccagagcaccacacccaagagaagatcattacagaaatcaaatgacac 1739
 Db 2646 agaccttaagagccagagcaccacacccaagagaagatcattacagaaatcaaatgacac 2705
 Qy 1740 aagccttggtctctgcccacagagtcacccaaggctgtgaacaaactctcctcttgcc 1799
 Db 2706 aagccttggtctctgcccacagagtcacccaaggctgtgaacaaactctcctcttgcc 2765
 Qy 1800 tccatccttaacagagtcctccacagagcaaaccttccagcccacagacacaaacttagc 1859
 Db 2766 tccatccttaacagagtcctccacagagcaaaccttccagcccacagacacaaacttagc 2825
 Qy 1860 acaaatgtctccaaactttcaacaactacattctgaagacctctcagcctctttagaccac 1919
 Db 2826 acaaatgtctccaaactttcaacaactacattctgaagacctctcagcctctttagaccac 2885
 Qy 1920 tctcggcacctctgaagcagacatctctggtcactggccctatagccgtacagactac 1979
 Db 2886 tctcggcacctctgaagcagacatctctggtcactggccctatagccgtacagactac 2945
 Qy 1980 agctggaaaaacagctctcgtcgcacccatcctgaaataactagttcctcaaatctcaacaga 2039
 Db 2946 agctggaaaaacagctctcgtcgcacccatcctgaaataactagttcctcaaatctcaacaga 3005

Qy 2040 agtgggcatcagcacagaaaggaaacccagtgattgtgagctaccactggattgatccc 2099
 Db 3006 agtgggcatcagcacagaaaggaaacccagtgattgtgagctaccactggattgatccc 3065
 Qy 2100 ttgaccagtgtaaccacatcagcaaaagaaatgaccacaagctttggcgtttacagcaga 2159
 Db 3066 ttgaccagtgtaaccacatcagcaaaagaaatgaccacaagctttggcgtttacagcaga 3125
 Qy 2160 gtacagcccagcttcaagttccctcctggaaacatctccttccccaacacagttgtttc 2219
 Db 3126 gtacagcccagcttcaagttccctcctggaaacatctccttccccaacacagttgtttc 3185
 Qy 2220 cacggctgaagacttggctcccaaatctgcacactttgtctgttcagagcagcacacagtc 2279
 Db 3186 cacggctgaagacttggctcccaaatctgcacactttgtctgttcagagcagcacacagtc 3245
 Qy 2280 accaacaacactgtcctctcctcagcctcagtcacaacagctgtgtgtgaaccccttgcctca 2339
 Db 3246 accaacaacactgtcctctcctcagcctcagtcacaacagctgtgtgtgaaccccttgcctca 3305
 Qy 2340 caatggcggaatgcgtcgcagacaacacacagcgtgtgctacacactgcaggtgccgccttc 2399
 Db 3306 caatggcggaatgcgtcgcagacaacacacagcgtgtgctacacactgcaggtgccgccttc 3365
 Qy 2400 ctggcaaggggatgatgcagtgaggatgaaatgagtgagtgctgcgaacccctgcacctc 2459
 Db 3366 ctggcaaggggatgatgcagtgaggatgaaatgagtgagtgctgcgaacccctgcacctc 3425
 Qy 2460 cacagccacgtgcacaatactcaggggataccttttatctgcaaaatgcgcggttggtgatacca 2519
 Db 3426 cacagccacgtgcacaatactcaggggataccttttatctgcaaaatgcgcggttggtgatacca 3485
 Qy 2520 gttggaaaaaggatatgcgaatttgggttagaaccttcgtgcagagttctaattaaaag 2579
 Db 3486 gttggaaaaaggatatgcgaatttgggttagaaccttcgtgcagagttctaattaaaag 3545
 Qy 2580 aactttctttaatacaactgtggaaaaacattcagacctacacaaagttgaaaaatgagat 2639
 Db 3546 aactttctttaatacaactgtggaaaaacattcagacctacacaaagttgaaaaatgagat 3605
 Qy 2640 caccaaaacgtttaaataatgtgttttcaagcgttcaactagttacatccgatctacagttca 2699
 Db 3606 caccaaaacgtttaaataatgtgttttcaagcgttcaactagttacatccgatctacagttca 3665
 Qy 2700 cgccttagggagtcaccaacgcggtgggtgatctcactgcacaacacttttccctggcctc 2759
 Db 3666 cgccttagggagtcaccaacgcggtgggtgatctcactgcacaacacttttccctggcctc 3725
 Qy 2760 caatgtgacgctatttgacctgggctgataggatgcagaaaatgtgtcaactcctcgaagtc 2819
 Db 3726 caatgtgacgctatttgacctgggctgataggatgcagaaaatgtgtcaactcctcgaagtc 3785
 Qy 2820 ctctcgtgaggtctcgcagctcttgggatactcagaggcgggatacttttagagcggcagctt 2879
 Db 3786 ctctcgtgaggtctcgcagctcttgggatactcagaggcgggatacttttagagcggcagctt 3845
 Qy 2880 gtgcgaagcgggaaggtcccggaatgtgcacaagacacccctccctcgtcactgcacctggacgg 2939
 Db 3846 gtgcgaagcgggaaggtcccggaatgtgcacaagacacccctccctcgtcactgcacctggacgg 3905
 Qy 2940 cgttgcctcgtgcagtgcaagtgcgggatactttcaagtccaacaaga tggaccactcctg 2999
 Db 3906 cgttgcctcgtgcagtgcaagtgcgggatactttcaagtccaacaaga tggaccactcctg 3965
 Qy 3000 ccgagcagtgtaagatggataggcttgaataatgaaacctgcagctgagttgccatttgg 3059
 Db 3966 ccgagcagtgtaagatggataggcttgaataatgaaacctgcagctgagttgccatttgg 4025
 Qy 3060 ccttgggtggtctcactgcagaaacccctatcagcttatcactgtggtgactgcagcgcgc 3119
 Db 4026 ccttgggtggtctcactgcagaaacccctatcagcttatcactgtggtgactgcagcgcgc 4085
 Qy 3120 gggagggtggggtcctcgtctca tccctagggcatcgcactgattgttaccctgttcgagaaaaga 3179

Qy	5340	accaatggccctttgttgctcacggctcgcaacctaaactgagagatttctgagctctgcga	5399
Db	6306	accaatggccctttgttgctcacggctcgcaacctaaactgagagatttctgagctctgcga	6365
Qy	5400	gtcactctgaagccacacagactaggcttcttgctccttcgcg	5441
Db	6366	gtcactctgaagccacacagactaggcttcttgctccttcgcg	6407
RESULT	5		
AAZ88007			
ID	AAZ88007	standard; cDNA; 6407 BP.	
XX	AC	AAZ88007;	
XX	XX	19-APR-2000 (first entry)	
XX	DT		
XX	XX	Rchd528 gene sequence SEQ ID NO:7.	
XX	XX	Cardiovascular disease; diagnosis; atherosclerosis; ischaemia; reperfusion; hypertension; restenosis; arterial inflammation; antiarteriosclerotic; vasotropic; hypotensive; ds.	
KW	KW	Homo sapiens.	
OS	XX	US6018025-A.	
PN	XX	25-JAN-2000.	
XX	PD		
XX	PF	06-OCT-1997; 97US-0944868.	
XX	XX	09-FEB-1996; 96US-0599654.	
PR	PR	10-FEB-1995; 95US-0386844.	
XX	PR	07-JUN-1995; 95US-0485573.	
XX	XX	(MILL-) MILLENIUM PHARM INC.	
PA	PA	(BGHM) BRIGHAM & WOMENS HOSPITAL.	
XX	XX	Falb DA, Gimbrone MA;	
PI	PI	WPI; 2000-136704/12.	
DR	DR	P-PSDB; AAY68447.	
DR	DR		
XX	XX	Isolated polypeptide for treating and diagnosing cardiovascular disease, such as, atherosclerosis, ischemia/reperfusion, hypertension, restenosis and arterial inflammation -	
PT	PT	Example; Fig 30; 122pp; English.	
XX	XX		
CC	CC	The present invention describes an isolated polypeptide (I) comprising either the amino acid sequence of 1481 residues, given in AAY68447, or an amino acid sequence encoded by the cDNA contained in plasmids pFCHP528A (ATCC 69985), pFCHP528B (ATCC 69986) and pFCHP528C (ATCC 69987). The polypeptide is useful in the treatment and diagnosis of cardiovascular disease, such as, atherosclerosis, ischemia/reperfusion, hypertension, restenosis and arterial inflammation. AAZ88001 to AAZ88040, and AAY68444 to AAY68457 represent sequences used in the exemplification of the present invention.	
CC	CC		
XX	XX	Sequence 6407 BP; 1594 A; 1762 C; 1520 G; 1509 T; 22 other;	
SQ			
Query Match 73.3%; Score 5098.4; DB 21; Length 6407;			
Best Local Similarity 94.3%; Pred. No. 0;			
Matches 5412; Conservative 21; Mismatches 8; Indels 301; Gaps 2;			
Qy	1	gttcgatgaagaattgccgcgttttcaacaaagatggaacgctcgagatgggaac	60
Db	666	gttcgatgaagaattgccgcgttttcaacaaagatggaacgctcgagatgggaac	725
Qy	61	agagagggcgatggggctgtcagaagaattggactgtgcacagccacagagccaccattc	120

Db 4026 ccttgggtggtcctaacgtggaaccctatcagcttatcactgtggtgatacgcagcgc 4085
QY 3120 gggagggtgggctcctgctccatctagcgatcgcaactgattgttacctggtgcagaagaa 3179
Db 4086 gggagggtgggctcctgctccatctcctagcgatcgcaactgattgttacctggtgcagaagaa 4145
QY 3180 taaaaatgacataagcaaaactcatttcaaaaagtggagatttccaaaatgtccccatatgc 3239
Db 4146 taaaaatgacataagcaaaactcatttcaaaaagtggagatttccaaaatgtccccatatgc 4205
QY 3240 tgaatacccaaaaaatcctcgtctcaacaagaattggggccgagaagctattgaaatgcataga 3299
Db 4206 tgaatacccaaaaaatcctcgtctcaacaagaattggggccgagaagctattgaaatgcataga 4265
QY 3300 gaattgggaagtaccaaaaaactcctccagatgacggatggtgtactactcgcctcaaatgt 3359
Db 4266 gaattgggaagtaccaaaaaactcctcccaaga tgcaggga tgtgtactactcgcctcaaatgt 4325
QY 3360 aaggaaatccagaacttgaacgaaacggaactctacccggcctacactggaactgccagatc 3419
Db 4326 aaggaaatccagaacttgaacgaaacggaactctacccggcctacactggaactgccagatc 4385
QY 3420 acggcatcttgaatttcccccggacagataataacccgtcttccatcagtgaataagcag 3479
Db 4386 acggcatcttgaatttcccccggacagataataacccgtcttccatcagtgaataagcag 4445
QY 3480 aagaagagactacttttaagtccaggagagagaggagactcatgctctgagccagtcacc 3539
Db 4446 aagaagagactacttttaagtccaggagagagaggagactcatgctctgagccagtcacc 4505
QY 3540 tgggacctctgctcagaggaccgcacacagagagctgcgccaggatttgcggagaccac 3599
Db 4506 tgggacctctgctcagaggaccgcacacagagagctgcgccaggatttgcggagaccac 4565
QY 3600 gctgagtcgcaagcaggaaaggagacagcatgcggggcgtgacccaagtcggaggagaca 3659
Db 4566 gctgagtcgcaagcaggaaaggagacagcatgcggggcgtgacccaagtcggaggagaca 4625
QY 3660 ggttggaatggaaccacagctgctcactcagcaccttctgtgtactgtgaacgtgaat 3719
Db 4626 ggttggaatggaaccacagctgctcactcagcaccttctgtgtactgtgaacgtgaat 4685
QY 3720 gtgggcccagtatcaagagatctctctgagtactgcacatggcactggcaccagggcg 3779
Db 4686 gtgggcccagtatcaagagatctctctgagtactgcacatggcactggcaccagggcg 4745
QY 3780 actattagccaggcgacacactagacttcagtcgacggagacctggtttcccttcgtttg 3839
Db 4746 actattagccaggcgacacactagacttcagtcgacggagacctggtttcccttcgtttg 4805
QY 3840 cactttagtaaaattgggtgggaggtttcccttttggatctgttttgagactgttccagaaa 3899
Db 4806 cactttagtaaaattgggtgggaggtttcccttttggatctgttttgagactgttccagaaa 4865
QY 3900 gaaggctctcttcccogagacacttccatagggcagcaatttggatgttcatttgcagcaa 3959
Db 4866 gaaggctctcttcccogagacacttccatagggcagcaatttggatgttcatttgcagcaa 4925
QY 3960 aatactggctgttaattatttccctgcccagcgctgctgctgctaaacaacagatagga 4019
Db 4926 aatactggctgttaattatttccctgcccagcgctgctgctgctaaacaacagatagga 4985
QY 4020 tgaagctaccacactgaagtcgaagatgtcgccattgaacggacagtggtttccatatgttt 4079
Db 4986 tgaagctaccacactgaagtcgaagatgtcgccattgaacggacagtggtttccatatgttt 5045
QY 4080 ctagggtgtcttatgtctacagtttccaagccagcccccacagtcgaggaatgtgtgaggc 4139
Db 5046 ctagggtgtcttatgtctacagtttccaagccagcccccacagtcgaggaatgtgtgaggc 5105
QY 4140 accgcacacaaactgcgaatgtgttttttaagtcaaggtgcacacatgtattttaagattttt 4199
.

Db 5106 acgcacacaaactgcaattgtgtttttaagtccaaggtgcacacatgtatttaagattttt 5165
QY 4200 tttaaaaatcctttgcagttaaaaatctcactttttcaacaagcctggatcgagcaaaac 4259
Db 5166 tttaaaaatccttctgcagttaaaaatctcacttttcaacaagcctggatcgagcaaaac 5225
QY 4260 aacttatatttgggttttagctggaggctcagcagcgagatgcaggcaggggggaacttt 4319
Db 5226 aacttatatttgggttttagctggaggctcagcagcgagatgcaggcaggggggaacttt 5285
QY 4320 tcatccatgggggcccagcctggggcctgggaactctcgcacccattgtgagggccaaggg 4379
Db 5286 tcatccatgggggcccagcctggggcctgggaactctcgcacccattgtgagggccaaggg 5345
QY 4380 cagctgcgtatgcaggagaaatgtcaaaactgaacgcagggtttccaccactcagaaagca 4439
Db 5346 camctgcgtatgcaggaggaatgtcaaaactgaacgcagggtttccaccactcagaaagca 5405
QY 4440 gcttgggtgagccccctgcagctggatgtgggttagagggtgggctgaatagcaggttaga 4499
Db 5406 gcttgggtgagccccctgcacactggatgtgggttagagggtgggctgaatagcaggttaga 5465
QY 4500 tttcctgcatacaacagtcgctttggggaagcgtgtggattcctcggggaagaacagggagcc 4559
Db 5466 tttcctgcatacaacagtcgctttggggaasctgtgtggattcctcggggaagaacagggagcc 5525
QY 4560 gagatggagccacacatgagtttgcctcacgggctactgcagcactttgtaccaccaactc 4619
Db 5526 gagatggagccacacatgaaattgctcacgggctactgcagcactttgtaccaccaactc 5585
QY 4620 catgtccacaaaaccccatgtaaaactttcaaacctcctcaaacctcctttatttcggctgaagaa 4679
Db 5586 catgtccacaaaaccccatgtaaaactttcaaacctcctcaaacctcctttatttcggctgaagaa 5645
QY 4680 ataaacttttttctcaccacagtcatttgttacctctcctcatgtgctgtgtgcgacccctcc 4739
Db 5646 ataaacttttttctcaccacagtcatttgttacctctcctcatgtgctgtgtgcgacccctcc 5705
QY 4740 agaaactgggttactctccagtcagtcgtgggagaaactgaagactccggttgcgtcgaggg 4799
Db 5706 agaaactgggttactctccagtcagtcgtgggagaaactgaagactccggttgcgtcgaggg 5765
QY 4800 aactgaggggttgaccttcggggaaggaagtctccactcatctttatttattatgcctgtgatg 4859
Db 5766 aactgaggggttgaccttcggggaaggaagtctccactcatctttatttattatgcctgtgatg 5825
QY 4860 tgggtcctgcaggagacatccagtcactcgggtgttctttaaattgccaccctggggaactgt 4919
Db 5826 tgggtcctgcaggagacatccagtcactcgggtgttctttaaattgccaccctggggaactgt 5885
QY 4920 gtttattgggctcttttggggcatcctcgggttttggatgaaagtgggggaatacacagaggt 4979
Db 5886 gtttattgggctcttttggggcatcctcgggttcctcgga tgaagtggagggaatacacagaggt 5945
QY 4980 aaagaaattgtctccacccctgaagcggggagtcgccgcttcacatttctggaaaatgggtcgag 5039
Db 5946 aaagaaattgtctccacccctgaagcggggagtcgccgcttcacatttctggaaaatgggtcgag 6005
QY 5040 ccactgggggacagttctgcgccgggcatggtgtgttttccaaggtcctcctcaaaataaatc 5099
Db 6006 ccactgggggacagttctgcgccgggcatggtgtgttttccaaggtcctcctcaaaataaatc 6065
QY 5100 cctattcttacaataactcttggccctgattggtttttaagcaagaactcctgtgtcccatgg 5159
Db 6066 cctattcttacaataactcttggccctgattggtttttaagcaagaactcctgtgtcccatgg 6125
QY 5160 tctccacacactccactcaccctgctgtagcaagagtccttagtcagggggaggtgcatttt 5219
Db 6126 tctccacacactccactcaccctgctgtagcaagagtccttagtcagggggaggtgcatttt 6185
QY 5220 agtagttaaatgcaacttatccatgagataaaaataaaaggaggaactgtttttatcagtcgga 5279
Db 6186 agtagttacattgcacttatccatgagataaaaataaaaggagagactgtttttatcagtcgga 6245
.

QY 5280 ggctaaacctaaatttcaaaagtgcgctcttttgaataatttggcctctctctgtaga 5339
 Db 6246 ggctaaacctaaatttcaaaagtgcgctcttttgaataatttggcctctctctgtaga 6305
 QY 5340 accaatggcccttctgtgctcagcgctcgcacctaactagagagttctgagctccctgca 5399
 Db 6306 accaatggcccttctgtgctcagcgctcgcacctaactagagagttctgagctccctgca 6365
 QY 5400 gctcaactgagccacagactaggtctcttggctctctccgc 5441
 Db 6366 gctcaactgagccacagactaggtctcttggctctctccgc 6407

RESULT 6

AAT36035
 ID AAT36035 standard; DNA; 6407 BP.

AC AAT36035;
 AC AAT36035;

XX 20-NOV-1996 (first entry)

DE rchd528 gene differentially expressed in cardiovascular disease.

XX Cardiovascular disease; differential expression; target gene;
 KW pathway gene; fingerprint gene; atherosclerosis; ischaemia;
 KW reperfusion; hypertension; restenosis; arterial inflammation;
 KW vector; antibody; diagnosis; gene therapy; drug screening;
 KW rchd528 gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT 19..4446
 FT CDS /*tag= a

XX W09624604-A1.

XX 15-AUG-1996.

XX 09-FEB-1996; 96WO-0501883.

XX 07-JUN-1995; 95US-0485573.

XX 10-FEB-1995; 95US-0386844.

XX (MILL-) MILLENNIUM PHARM INC.

XX Falb DA;

XX WPI; 1996-384391/38.
 XX P-PSDB; AAW03740.

XX New genes differentially expressed in cardiovascular disease - and
 PT related vectors, host cells, proteins and antibodies, for diagnosis,
 PT monitoring, treatment and drug screening

XX Claim 1; Fig 31A-E; 200pp; English.

XX 4 Novel genes, rchd502 (AAT36033), rchd523 (AAT36034), rchd528
 CC (AAT36035) and rchd534 (AAT36036), are differentially expressed in
 CC endothelial cells subjected to shear stress. HUVEC cells were exposed
 CC to laminar shear stress, which is thought to be responsible for the
 CC prevalence of atherosclerotic lesions in areas of unusual
 CC circulatory flow. RNAs from treated and control cells were used to
 CC generate cDNA libraries. Differentially expressed bands were
 CC identified by electrophoresis, subcloned and sequenced. The rchd528
 CC gene product (AAW03740) was also sequenced. rchd528 is very highly
 CC expressed in the heart. 4 other genes (see also AAT36033-36) are
 CC differentially expressed in endothelial cells exposed to interleukin-1.
 CC Detection of these 8 novel genes in excess of normal levels allows the
 CC diagnosis of cardiovascular diseases (CVD). The genes can be used to
 CC generate diagnostic probes, to produce recombinant gene products, to
 CC breed transgenic animal models of CVD, in gene replacement therapy,

CC or (as antisense, ribozyme or triplex sequences) to treat CVD.
 XX Sequence 6407 BP; 1592 A; 1763 C; 1523 G; 1507 T; 22 other;
 SQ Matches 5377; Conservative 21; Mismatches 43; Indels 301; Gaps 2;
 Query Match 72.5%; Score 5042.4; DB 17; Length 6407;
 Best Local Similarity 93.6%; Pred No. 0;
 Matches 5377; Conservative 21; Mismatches 43; Indels 301; Gaps 2;
 QY 1 gttcgtgaaagaattcccgctcttttcaacaaagaaggagagcctcctcgagatgggaac 60
 Db 666 gttcgtgaaagaattcccgctcttttcaacaaagaaggagagcctcctcgagatgggaac 725
 QY 61 agagaggcgatgggggtgctcagaagaatggactgtgcacagccaaagaggccacacttc 120
 Db 726 agagaggcgatgggggtgctcagaagaatggactgtgcacagccaaagaggccacacttc 785
 QY 121 ggtctggagcctctcttctag 180
 Db 786 ggtctggagcctctcttctag 845
 QY 181 gaagagaaattctctcaggaccagatctctctctctctctctctctctctctctctctctc 240
 Db 846 gaagagaaattctctcaggaccagatctctctctctctctctctctctctctctctctc 905
 QY 241 ctc 300
 Db 906 ctc 965
 QY 301 tggcctcagagctcctcagtcagtcacaaacaaagaacatgcatgttctactccgtgtctac 360
 Db 966 tggcctcagagctcctcagtcagtcacaaacaaagaacatgcatgttctactccgtgtctac 1025
 QY 361 tgatggtggcccgagaaacgctgcgactcttggacggtcagtcgtggacgtgtgacgaagac 420
 Db 1026 tgatggtggcccgagaaacgctgcgactcttggacggtcagtcgtggacgtgtgacgaagac 1085
 QY 421 agaaagcttcccaagactccagaattgccagacttcacacagtcctcagtcctctctcacc 480
 Db 1086 agaaagcttcccaagactccagaattgccagacttcacacagtcctcagtcctctctcacc 1145
 QY 481 ctctgcagtggaatcgagaagaacagtagtaactgggaatccacagggggatgggaatt 540
 Db 1146 ctctgcagtggaatcgagaagaacagtagtaactgggaatccacagggggatgggaatt 1205
 QY 541 cattgaaccatccacagaaatgaattggacttacgtctcttgcg-tggcaaatgattc 599
 Db 1206 cattgaaccatccacagaaatgaattggacttacgtctcttgcg-tggcaaatgattc 1265
 QY 600 cccaacctttggagaacatcagcttgcagcagctctcaggtgcacaaatggaagtcccat 659
 Db 1266 cccaacctttggagaacatcagcttgcagcagctctcaggtgcacaaatggaagtcccat 1325
 QY 660 gtctcagactgagactgtgtcagtcagtcgaccccatgagaggtggagagatcaactgc 719
 Db 1326 gtctcagactgagactgtgtcagtcagtcgaccccatgagaggtggagagatcaactgc 1385
 QY 720 acactggtcttgaccacacagcacaacatctgcagatgtgacaggaagctctgttccata 779
 Db 1386 acactggtcttgaccacacagcacaacatctgcagatgtgacaggaagctctgttccata 1445
 QY 780 tcttgaaggtgtaagtctcagtcggtgacccagttctcagactctcagactctgtacagtc 839
 Db 1446 tcttgaaggtgtaagtctcagtcggtgacccagttctcagactctcagactctgtacagtc 1505
 QY 840 aggaagtccacacagcattggagatagaggtatttcagagtccttcacatctctctc 899
 Db 1506 aggaagtccacacagcattggagatagaggtatttcagagtccttcacatctctc 1565
 QY 900 ggaagcttgaattcattcagcaccacacgctgagaaacgttcaa----- 940
 Db 1566 ggaagcttgaattcattcagcaccacacgctgagaaacgttcaa----- 1625

QY 941 ----- 940
Db 1626 gccaggccaagcactaggtgacagttccgcgaatgcagagagacaggactctctgggtgcc 1685
QY 941 ----- 940
Db 1686 ctctctcggccaccacaccttggtactgtcactggaacggggaacgcacactgcggtc 1745
QY 941 ----- 940
Db 1746 tgtcacccctcaccaaacaccagatgagcacgactctctggggaagcgagccctgcagc 1805
QY 941 ----- 940
Db 1806 ggccatgcccgaagaaacagagggtgctctctgcacgtaaacgtgacgcagacatggg 1865
QY 941 ----- 940
Db 1866 cctggtctcaggtcactggccgctccagtgcaactcggagtcgctgggattagctacgg 1925
QY 960 tcaagtgcgtggtgcacagctattgaacaaaaggacttccagcgaccacacagaccaccta 1019
Db 1926 tcaagtgcgtggtgcacagctattgaacaaaaggacttccagcgaccacacagaccaccta 1985
QY 1020 cctgtcatctactttccaccaaaaggaaacggcggttactgtccattacagataaacagttc 1079
Db 1986 cctgtcatctactttccaccaaaaggaaacggcggttactgtccattacagataaacagttc 2045
QY 1080 atcctcagaacattgtggagagctcaactcttattattataaaatctcaaacctctccaatc 1139
Db 2046 atcctcagaacattgtggagagctcaactcttattattataaaatctcaaacctctccaatc 2105
QY 1140 agagtattcctcttctctatgctcagactgagagaagtaacatctcatccttatgacgg 1199
Db 2106 agagtattcctcttctctatgctcagactgagagaagtaacatctcatccttatgacgg 2165
QY 1200 ggaatatgctcagcctctactgagtcgccagttctgcatacatccaacctctcgtccta 1259
Db 2166 ggaatatgctcagcctctactgagtcgccagttctgcatacatccaacctctcgtccta 2225
QY 1260 cacaccacacataataatgcgaacacttcgggtgttctcgtggaaactgatctgagtttgt 1319
Db 2226 cacaccacacataataatgcgaacacttcgggtgttctcgtggaaactgatctgagtttgt 2285
QY 1320 tagtgactctcctcctctctctcctcctctctctctctctctctctctctctctctctct 1379
Db 2286 tagtgactctcctcctcctctctcctcctcctcctcctcctcctcctcctcctcctct 2345
QY 1380 gcctcgcctctcgtgtcaaatcccaaccatttatttttcaatcaattttaccatacaaccag 1439
Db 2346 gcctcgcctctcgtgtcaaaatcccaaccatttatttttcaatcaattttaccatacaaccag 2405
QY 1440 ggcctctgtcatctactataaagtctacctctgctgatccacaccatggcttctcctcaacc 1499
Db 2406 ggcctctgtgtcatctactataaagtctacctctgctgatccacaccatggcttctcctcaacc 2465
QY 1500 atcacctttaccagttaccttaacgacatctacatctgcgccacttctctgtctcaaaaac 1559
Db 2466 atcacctttaccagttaccttaacgacatctacatctgcgccacttctctgtctcaaaaac 2525
QY 1560 aacettggccacagtatctcttccctcctgctcctcctcctcctcctcctcctcctcctgac 1619
Db 2526 aacettggccacagtatctctcctcctcctcctcctcctcctcctcctcctcctcctcct 2585
QY 1620 ttcatattcagacatcaacaatgacatcatctcatgacaatgtctccatagtagtcaaaactgc 1679
Db 2586 ttcatattcagacatcaacaatgacatcatctcatgacaatgtctccatagtagtcaaaactgc 2645
QY 1680 agaccttaagagccagagcccccacacccaagagaaagtcatcacgaaatcaaaagtcaacc 1739
Db 2646 agaccttaagagccagagacccccacacccaagagaaagtcatcacgaaatcaaaagtcaacc 2705
QY 1740 aagcctgggtgtctctgcccacagagtccacccaagctgtaacaaacaaactctcctttgcc 1799

Db 2706 aagcctgggtgtctctgcccacagagtcacccaagctgttaacaaacaaactctctttgcc 2765
QY 1800 tccatctctaacagagtcctctccacagagcaaacctctccagccaagcaccacttagc 1859
Db 2766 tccatctctaacagagtcctctccacagagcaaacctctccagccaagcaccacttagc 2825
QY 1860 acaaatctctccaaactttcaaaactacattctgaagacctctcagctctcttagaccac 1919
Db 2826 acaaatctctccaaactttcaaaactacattctgaagacctctcagctctcttagaccac 2885
QY 1920 tcttggcacctgtcaagcacagcatctctggtcactggccctatagccgttacagactac 1979
Db 2886 tcttggcacctgtcaagcacagcatctctggtcactggccctatagccgttacagactac 2945
QY 1980 agctggaaaaacagctctcgtctgacctatcctgaaataactagttctcaaatctcaacaga 2039
Db 2946 agctggaaaaacagctctcgtctgacctatcctgaaataactagttctcaaatctcaacaga 3005
QY 2040 aggtggcatcagcacagaaaaggaaacccagtgattgtgtgactaccactggattgatccc 2099
Db 3006 aggtggcatcagcacagaaaaggaaacccagtgattgtgtgactaccactggattgatccc 3065
QY 2100 tttagccagtgtagcccaatcagcaaaagaaatgaccacaagaacttggtgttacagcaga 2159
Db 3066 tttagccagtgtagcccaatcagcaaaagaaatgaccacaagaacttggtgtgttacagcaga 3125
QY 2160 gtacagcccagcttcaagttccctcgggaacatctccttctcccaacccacagttgtttc 2219
Db 3126 gtacagcccagcttcaagttccctcgggaacatctccttctcccaacccacagttgtttc 3185
QY 2220 cacggtggaagacttggtctccaaactgcacactttgtgttcagagcagcacacagtc 2279
Db 3186 cacggtggaagacttggtctccaaactgcacactttgtgttcagagcagcacacagtc 3245
QY 2280 accaacaacactgtcctcttcagcctcagtcacacagctgtgtgtgtgaacccctgtcttca 2339
Db 3246 accaacaacactgtcctcttcagcctcagtcacacagctgtgtgtgtgaacccctgtcttca 3305
QY 2340 caatggcgaaatgggtcgcagacaacacccgcgtgtgtaccactgcagggtgccgccttc 2399
Db 3306 caatggcgaaatgggtcgcagacaacacccgcgtgtgtaccactgcagggtgccgccttc 3365
QY 2400 ctggcaaggagtgattgagtggtggaatgagtgccgtgcgaacccctgcccatc 2459
Db 3366 ctggcaaggagtgattgagtggtggaatgagtgccgtgcgaacccctgcccatc 3425
QY 2460 cacagccactgtcaacaataactcagggaactcttatctgcaaaatgcocgggtgggtacca 2519
Db 3426 cacagccactgtcaacaataactcagggaactcttatctgcaaaatgcocgggtgggtacca 3485
QY 2520 gttggaaaaaggagatgtgcaatttgggttagaacctctcgtacagagtttaaatlaaagag 2579
Db 3486 gttggaaaaaggagatgtgcaatttgggttagaacctctcgtacagagtttaaatlaaagag 3545
QY 2580 aactttctttaaatacaactgtggaataacattcagactcaagaagttgaaatagat 2639
Db 3546 aactttctttaaatacaactgtggaataacattcagactcaagaagttgaaatagat 3605
QY 2640 caccaaaacgttaaatatgtgttttccagcgtttaccctagttacatccgatctacagttca 2699
Db 3606 caccaaaacgttaaatatgtgttttccagcgtttaccctagttacatccgatctacagttca 3665
QY 2700 cgcctctaggaggtccaacgcgggtgtgtactcactcaaacacactttccctcggcctc 2759
Db 3666 cgcctctaggaggtccaacgcgggtgtgtactcactcaaacacactttccctcggcctc 3725
QY 2760 caatgtgacgctatttgacctggctgagtagtgagaaatgtgtcaactcctcgaagtc 2819
Db 3726 caatgtgacgctatttgacctggctgagtagtgagaaatgtgtcaactcctcgaagtc 3785
QY 2820 ctctgtcagggctgcacagctctcttgggatctcagagggcagatcttttagagcggcagctt 2879

Db	3786	ctctgtctgaagttctgcagctctcttggtatctcagaggcggaatcttttagagcggcggaagctt	3845
Qy	2880	gtgcagcggaaagatcccgaaatgtgacaaagacacctcccattctgcactgcactgcactggacgg	2939
Db	3846	gtgcagcggaaagatcccgaaatgtgacaaagacacctcccattctgcactgcactggacgg	3905
Qy	2940	cgttgccctgtgccagtgcgaatcgggatactttcagttcaacaagatggacaactctgtg	2999
Db	3906	cgttgccctgtgccagtgcgaatcgggatactttcagttcaacaagatggacaactctgtg	3965
Qy	3000	ccagcatgtgaagatggatagacttgaaatgaaacctgcatggttgcacattctgg	3059
Db	3966	ccagcatgtgaagatggatagacttgaaatgaaacctgcatggttgcacattctgg	4025
Qy	3060	ccttggtggtctcaactgtggaacccctatcagcttatactgtggtgatacgcagcgc	3119
Db	4026	ccttggtggtctcaactgtggaacccctatcagcttatactgtggtgatacgcagcgc	4085
Qy	3120	ggaggtgggtcctctcatcctctggaatgcagctggtatgttaccctgttgcagaagaa	3179
Db	4086	ggaggtgggtcctctcatcctctggaatgcagctggtatgttaccctgttgcagaagaa	4145
Qy	3180	taaaaatgcataaagcaaaactcatctcaaaagtggagattccaaatgtcccattatgc	3239
Db	4146	taaaaatgcataaagcaaaactcatctcaaaagtggagattccaaatgtcccattatgc	4205
Qy	3240	tgaatcccccaaaaatcctcgctcaagaagtgggcccgaagaactattgaaatgcata	3299
Db	4206	tgaatcccccaaaaatcctcgctcaagaagtgggcccgaagaactattgaaatgcata	4265
Qy	3300	gaatgaagtaccaaaaaactcctccagatgaaggatgtgtactactgcgtacaaagctg	3359
Db	4266	gaatgaagtaccaaaaaactcctccagatgaaggatgtgtactactgcgtacaaagctg	4325
Qy	3360	aaggaaatccagaacttgaacgaacacggactatacccgccctacactggactgccagatc	3419
Db	4326	aaggaaatccagaacttgaacgaacacggactatacccgccctacactggactgccagatc	4385
Qy	3420	acggactcttgcattttcccgcgacagtataaacccgtcttctcatcagtgatgaagcag	3479
Db	4386	acggactcttgcattttcccgcgacagtataaacccgtcttctcatcagtgatgaagcag	4445
Qy	3480	aagaagagactacttttaagtccaggagagagaggactcatctgtcttgagcagtcacc	3539
Db	4446	aagaagagactacttttaagtccaggagagagaggactcatctgtcttgagcagtcacc	4505
Qy	3540	tgggacctctgctcagagagaccgcacaggagctgcgccagatctgtcggagagccac	3599
Db	4506	tgggacctctgctcagagagaccgcacaggagctgcgccagatctgtcggagagccac	4565
Qy	3600	gctgagtggcaagcaggaagggacagcagtcggcgctgaccacagtgaggagagaca	3659
Db	4566	gctgagtggcaagcaggaagggacagcagtcggcgctgaccacagtgaggagagaca	4625
Qy	3660	ggtgagatggaaaccacaggtgcttatttcagacaccttctgttactgtgaacgtgaat	3719
Db	4626	ggtgagatggaaaccacaggtgcttatttcagacaccttctgttactgtgaacgtgaat	4685
Qy	3720	gtgggcccagatcaagagatctctctgagtgcacctgcacatggcactggcaccaggcg	3779
Db	4686	gtgggcccagatcaagagatctctctgagtgcacctgcacatggcactggcaccaggcg	4745
Qy	3780	actatagccagggcagacacactagacttcagtgaggagcctggttttcccttcgtttg	3839
Db	4746	actatagccagggcagacacactagacttcagtgaggagcctggttttcccttcgtttg	4805
Qy	3840	cactttagtaaatgtggtgggaggttcccttttgatctgttttgagactgttccagaaa	3899
Db	4806	cactttagtaaatgtggtgggaggttcccttttgatctgttttgagactgttccagaaa	4865
Qy	3900	gaaggttctcttcccgagacacttccataggcagcaaatcttggtgattcatcttgacgaa	3959
Db	4866	gaaggttctcttcccgagacacttccataggcagcaaatcttggtgattcatcttgacgaa	4925

Qy 3599 cgctgagtgcaagcaggagagacaggcatgcggggcgtagccacagtgaggagac 3658
Dbb 901 cgctgagtgcaagcaggagagacaggcatgcggggcgtagccacagtgaggagac 960
Qy 3659 aggtgagtgtagaacacacagctgctcatcagcacctttgttactgtgaacgtgaa 3718
Dbb 961 aggtgagtgtagaacacacagctgctcatcagcacctttgttactgtgaacgtgaa 1020
Qy 3719 tgtgggcccagttatcaagagagctctctctgagtgactgcacacatggcactggccaccaggc 3778
Dbb 1021 tgtgggcccagttatcaagagagctctctctgagtgactgcacacatggcactggccaccaggc 1080
Qy 3779 gactattagccaggcagacacactagacttcagtgccaggagcctgttttcccttcggtt 3838
Dbb 1081 gactattagccaggcagacacactagacttcagtgccaggagcctgttttcccttcggtt 1140
Qy 3839 gcactttagtaaatgggtgggaggtttccctttggatctctgttttgagactgttcagaa 3898
Dbb 1141 gcactttagtaaatgggtgggaggtttccctttggatctctgttttgagactgttcagaa 1200
Qy 3899 agaaggcttcttcccgagacacttccatagggagcaatttggtagtctatttgcagca 3958
Dbb 1201 agaaggcttcttcccgagacacttccatagggagcaatttggtagtctatttgcagca 1260
Qy 3959 aaatactggctgttaattttctctgccagcgctgcgtgctaaacacacagatgagg 4018
Dbb 1261 aaatactggctgttaattttctctgccagcgctgcgtgctaaacacacagatgagg 1320
Qy 4019 atgagctacacactgaagctcgaagatctgccattgaacggacagtgctttcatatgtt 4078
Dbb 1321 atgagctacacactgaagctcgaagatctgccattgaacggacagtgctttcatatgtt 1380
Qy 4079 tctaggttcttatgtctacagtttccaagccagcccccacagtgagaaatgtgtgagg 4138
Dbb 1381 tctaggttcttatgtctacagtttccaagccagcccccacagtgagaaatgtgtgagg 1440
Qy 4139 caccgcacacactgcaatgtgtttttaaagtcgaagtgacacatgtattttaaagattttt 4198
Dbb 1441 caccgcgcacacactgcaatgtgtttttaaagtcgaagtgacacatgtattttaaagattttt 1500
Qy 4199 ttttaaaatctctttgcagtttaaatctcaacttcttcaaacagcctggatcagggcaaaa 4258
Dbb 1501 tttttaaactctttgcagtttaaatctcaacttcttcaaacagcctggatcagggcaaaa 1560
Qy 4259 caacttatattgttttagctgaggtctcagcagccagtgatgcaggcaggggggcactt 4318
Dbb 1561 caacttatattgttttagctgaggtctcagcagccagtgatgcaggcaggggggcactt 1620
Qy 4319 ttcatccatgaggcccgagcctgggctgggactctgatcaccattgtggaggccagag 4378
Dbb 1621 ttcatccatgaggcccgagcctgggctgggactctgatcaccattgtggaggccagag 1679
Qy 4379 gcagctcgctatgaggagaaatgtcaaacgtgaacgcaggtttcaccactctaggaaagc 4438
Dbb 1680 gcagctcgctatgaggagaaatgtcaaacgtgaacgcaggtttcaccactctaggaaagc 1739
Qy 4439 agcttttagccctgcagctgagtggttagaggagatgggctgaatagccaggttag 4498
Dbb 1740 agcttttagccctgcagctgagtggttagaggagatgggctgaatagccaggttag 1799
Qy 4499 atttccctgatacaacagtgctttgggaagctgtgtggatttccctgaggaagacagggagc 4558
Dbb 1800 atttccctgatacaacagtgctttgggaagctgtgtggatttccctgaggaagacagggagc 1859
Qy 4559 cgagatggagccacacatgagttgtctcacgcggtactgcagcactttgttaccagaaatc 4618
Dbb 1860 cgagatggagccacacatgagttgtctcacgcggtactgcagcactttgttaccagaaatc 1919
Qy 4619 tcatgtccaaaaacccatgtaaactttcaaccactcaaaagctgtttattcggctgaaga 4678
Dbb 1920 tcatgtccaaaaacccatgtaaactttcaaccactcaaaagctgtttattcggctgaaga 1979
Qy 4679 aataacttttttctcaaccagtcatttgttaacttttcatatggtgtgtgcaccctc 4738

Dbb 1980 aataacttttttctcaaccagtcatttgttaacttttcatatggtctatgtgcaccctc 2039
Qy 4739 cagaacactgggttatacttccagtcagtggtggagaactgaagaacttcggttggctcgag 4798
Dbb 2040 cagaacactgggttatacttccagtcagtggtggagaactgaagaacttcggttggctcgag 2099
Qy 4799 gaactgaggggttgacacttcgggaagagttccactcatcttatttatttattcctgtgat 4858
Dbb 2100 gaactgaggggttgacacttcgggaagagttccactcatcttatttatttattcctgtgat 2159
Qy 4859 gtgggtcctgccaggagagacatccagttactcgtgtctttaaattgcccacttgggaactg 4918
Dbb 2160 gtgggtcctgccaggagagacatccagttactcgtgtctttaaattgcccacttgggaactg 2219
Qy 4919 tgtttattggccttcttggggcatcctcgtgttttggatgaagtgggggaaatcacagagt 4978
Dbb 2220 tgtttattggccttcttggggcatcctcgtgttttggatgaagtgggggaaatcacagagt 2279
Qy 4979 aaaaagaattgtctccacactgaagcgggagtcctcgttccacttccatttctgaaatggtgca 5038
Dbb 2280 aaaaagaattgtctccacactgaagcgggagtcctcgttccacttccatttctgaaatggtgca 2339
Qy 5039 gccactggggacagttctgcccccgggcatggttcttccaaagtcctcttaataataat 5098
Dbb 2340 gccactggggacagttctgcccccgggcatggttcttccaaagtcctcttaataataat 2399
Qy 5099 cccatttttaacataaactcctggccctgatggtttttaagcaagaactcctgtgtcccatg 5158
Dbb 2400 cccatttttaacataaactcctggccctgatggtttttaagcaagaactcctgtgtcccatg 2459
Qy 5159 gtctccacactcacatcacactcgtgtgacaaagtcctcagtagtcagggaggtgcattt 5218
Dbb 2460 gtctccacactcacatcacactcgtgtgacaaagtcctcagtagtcagggaggtgcattt 2519
Qy 5219 tagtagttaaattgcacttatccatgagataaaataaaaggagaactgttttatcagtg 5278
Dbb 2520 tagtagttaaattgcacttatccatgagataaaataaaaggagaactgttttatcagtg 2579
Qy 5279 aggttaacttaaaatttcaaaagtcgcctttttgaaatctctggcctctctctctgtag 5338
Dbb 2580 aggttaacttaaaatttcaaaagtcgcctttttgaaatctctggcctctctctctgtag 2639
Qy 5339 aaccaatggccttctgtgcacgcctcgcacttaactgaagagttctgagcctctg 5398
Dbb 2640 aaccaatggccttctgtgcacgcctcgcacttaactgaagagttctgagcctctg 2699
Qy 5399 agctcacctgagccacagactaggtcttctgtccttccgcagcagcctgcctcaccc 5458
Dbb 2700 agctcacctgagccacagactaggtcttctgtccttccgcagcagcctgcctcaccc 2759
Qy 5459 ccagaacccgcagctgttggaagagccatgtaggaggctaattcccaggcatataactcc 5518
Dbb 2760 ccagaacccgcagctgttggaagagccatgtaggaggctaattcccaggcatataactcc 2819
Qy 5519 actgcttcagctgacgtcacagctgcacaaatcactcctctatcggagccagaactt 5578
Dbb 2820 actgcttcagctgacgtcacagctgcacaaatcactcctctatcggagccagaactt 2879
Qy 5579 cagctcccaaaaataagtggttctcctgaaacacattcttgggagaaatcccaactcg 5638
Dbb 2880 cagctcccaaaaataagtggttctcctgaaacacattcttgggagaaatcccaactcg 2939
Qy 5639 agaaaacgggtctctgtgagttcccaacatgcttctgttcaatgggttcttccgtatgg 5698
Dbb 2940 agaaaacgggtctctgtgagttcccaacatgcttctgttcaatgggttcttccgtatgg 2999
Qy 5699 agtgagtaagaagtggtttattttgttcttaactggaataaaggagcagccacaaa 5758
Dbb 3000 agtgagtaagaagtggtttattttgttcttaactggaataaaggagcagccacaaa 3059
Qy 5759 ggttgaggtcacagctccacagtttccagggcggttgggggagggagcact 5818

Db 3060 ggttgaggtcacacagtcctccacagtttccagagagcgttttgggggtggggagggcaccct 3119
Qy 5819 ccagagcatgaggtctcaaggagggacatagtaaacatgtctgtgacccagtgagggaagg 5878
Db 3120 ccagagcatgaggtctcaaggagggacatagtaaacatgtctgtgacccagtgagggaagg 3179
Qy 5879 gagaggccagctgcactcctctcacaggggttctctagctgcagaaagggtcccgccctagggccg 5938
Db 3180 gagaggccagctgcactcctctcacaggggttctctagctgcagaaagggtcccgccctagggccg 3239
Qy 5939 aggggaaacacctgatagcagaaaggccctggatgcacacctggcagccgaggtctctcc 5998
Db 3240 aggggaaacacctgatagcagaaaggccctggatgcacacctggcagccgaggtctctcc 3299
Qy 5999 gccagacacagtgctccatctcagccctgcacctggaggtggtgtgtattcaactgcagcag 6058
Db 3300 gccagacacagtgctccatctcagccctgcacctggaggtggtgtgtattcaactgcagcag 3359
Qy 6059 atgcacaatcctgcacaaatccccacagatggggagagtgtagagaaaggggcgaagtg 6118
Db 3360 atgcacaatcctgcacaaatccccacagatggggagagtgtagagaaaggggcgaagtg 3419
Qy 6119 atgtgtactgtcaagagatgcttaaacctccatagagagagccgggaggggagccat 6178
Db 3420 atgtgtactgtcaagagatgcttaaacctccatagagagagccgggaggggagccat 3479
Qy 6179 ctgtgtgtcccgctcacacactgcagcagggagaggtggtgctggctgctccctggcatcag 6238
Db 3480 ctgtgtgtcccgctcacacactgcagcagggagaggtggtgctggctgctccctggcatcag 3539
Qy 6239 tggtttgggttaagctccagagggctctattggccattgtcttttctctctgccccttgagc 6298
Db 3540 tggtttgggttaagctccagagggctctattggccattgtcttttctctctgccccttgagc 3599
Qy 6299 cagcttaagccctggagctgttttcttaggcgagatgaactgacatgctctcaccatga 6358
Db 3600 cagcttaagccctggagctgttttcttaggcgagatgaactgacatgctctcaccatga 3659
Qy 6359 ccaggctctggcgaagctctcctcacagtattcttgagaggtgggcatggagtgcccatt 6418
Db 3660 ccaggctctggcgaagctctcctcacagtattcttgagaggtgggcatggagtgcccatt 3719
Qy 6419 tctcaggtacagaaaccttcagagagataaaatgcttgccctgtagaagcagagactgaa 6478
Db 3720 tctcaggtacagaaaccttcagagagataaaatgcttgccctgtagaagcagagactgaa 3779
Qy 6479 accctgtccgctgactccccagctactctgcccactgtagcccctgctcctactgtc 6538
Db 3780 accctgtccgctgactccccagctactctgcccactgtagcccctgctcctactgtc 3839
Qy 6539 ctggcacacccctcaccatcctgtatataccttaataatacaagaggggcaagagagaaagg 6598
Db 3840 ctggcacacccctcaccatcctgtatataccttaataatacaagaggggcaagagagaaagg 3899
Qy 6599 ctttaagataagttatttttttaagaaaccttaataatttttttaagaaagtaacccaaat 6658
Db 3900 ctttaagataagttatttttttaagaaaccttaataatttttttaagaaagtaacccaaat 3959
Qy 6659 tagtgacgtgaaatgc 6674
Db 3960 tagtgacgtgaaatgc 3975

RESULT 8
AAS02424
ID AAS02424 standard; cDNA; 3116 BP.
XX
AC AAS02424;
XX AC
DT 18-JUL-2001 (first entry)
XX
DE Human secreted protein, cDNA #30.
XX

KW Human; secreted protein; immunogen; antibody; diagnosis;
KW rheumatoid arthritis; hyperproliferative disorder; neoplasm;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; Alzheimer's disease; bacterial infection; viral infection;
KW fungal infection; corneal infection; wound healing; cell culture;
KW epithelial cell proliferation; skin ageing; transplantation;
KW tissue regeneration; chemotaxis; food additive; ss.
XX Homo sapiens.
XX
XX WO200123546-A1.
XX PD 05-APR-2001.
XX 26-SEP-2000; 2000WO-US26323.
XX 27-SEP-1999; 99US-0155805.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA Rosen CA, Ruben S, Komatsoulis GA;
XX PI
XX WPI; 2001-266150/27.
XX DR P-PSDB; AAU01751.
XX Nucleic acids encoding 37 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
PT
XX Disclosure; Page 432-433; 494pp; English.
XX The sequence encodes a human secreted protein of the invention. The
XX polynucleotides, polypeptides and antibodies raised against them are used
XX to prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. The
XX polynucleotides and antibodies are also used in diagnosing a pathological
XX condition or susceptibility to a pathological condition. The antibodies
XX can also be used in alleviating symptoms associated with the disorders
XX and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX immunosorbent assays (ELISA). Disorders which are diagnosed or treated
XX include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX angioneurosis, nervous system disorders e.g. Alzheimer's disease,
XX infections caused by bacteria, viruses and fungi and ocular disorders
XX e.g. corneal infection. The polypeptides can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities.
SQ Sequence 3116 BP; 770 A; 743 C; 810 G; 793 T; 0 other;

Query Match 44.5%; Score 3092.8; DB 22; Length 3116;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3108; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
Qy 3577 gccacagatttctcgggagccacgctgagtgccaagcaggagggagcgatgcggg 3636
Db 2 gcacgagatttctcgggagccacgctgagtgccaagcaggagggagcgatgcggg 61
Qy 3637 gcgtgaccacagtgaggagacagtgatgtggaacacacaggtgctcattcacgacct 3696
Db 62 gcgtgaccacagtgaggagagcaggtggtgtaaacacacaggtgctcattcacgacct 121
Qy 3697 ttgtgttactgtgaacgtgaatgtggccagatccaagagagtgctctctgagtactgc 3756
Db 122 ttgtgttactgtgaacgtgaatgtggccagatccaagagagtgctctctgagtactgc 181
Qy 3757 accatggcaactggcaccagggcgactattagccaggggcagaccactagacttcagtcag 3816
|||||

Db 182 accatggcactggcaccaggcggaactatttagccaggggcagaccactagacttccagtgcag 241
QY 3817 ggaacctgtttcccttcctgtttccacattagtaataattgggtggagggtttcccttttggat 3876
Db 242 ggaacctgggtttcccttcctgtttccacatttagtaataattgggtggagggtttcccttttggat 301
QY 3877 ctgttttgagactgtttccagaagaagggttcctttcccgagacacttccatagggacga 3936
Db 302 ctgttttgagactgtttccagaagaagggttcctttcccgagacacttccatagggacga 361
QY 3937 atttgggtatttcatttgcagcaaaaatactggcttggtttaattatttccctgcacagcgct 3996
Db 362 atttgggtatttcatttgcagcaaaaatactggcttggtttaattatttccctgcacagcgct 421
QY 3997 gctgtgtaaaacacagatgaggatgagcgtaccactgaactgtgaagatgtccgccattga 4056
Db 422 gcgtgtaaaacacagatgaggatgagcgtaccactgaactgtgaagatgtccgccattga 481
QY 4057 acggacagtgttttcataatgtttctaggtgtcttatgtcacagtgttccaagccagccccc 4116
Db 482 acggacagtgttttcataatgtttctaggtgtcttatgtcacagtgttccaagccagccccc 541
QY 4117 cacagtggagaaatgtgtgagggcaccgcgcacacactgcaatgtgttttttaagtcaaggt 4176
Db 542 cacagtggagaaatgtgtgagggcaccgcgcacacactgcaatgtgttttttaagtcaaggt 601
QY 4177 gacacatgtatttaagaatttttttttaaaatctctttgcaagttaaatctcacatttttcaa 4236
Db 602 gacacatgtatttaagaatttttttttaaaatctctttgcaagttaaatctcacatttttcaa 661
QY 4237 acaagcctggatcagggcacaaacaaacttatatttgggttttttagctgaggtctcagcaggca 4296
Db 662 acaagcctggatcagggcacaaacaaacttatatttgggttttttagctgaggtctcagcaggca 721
QY 4297 gattgagcagggggggcaacttttccatcattgagggggccacgcctgggcttgggaactctg 4356
Db 722 gattgagcagggggggcaacttttccatcattgagggggccacgcctgggcttgggaactc-g 780
QY 4357 atcacatgttgagggccagagcagctgcgtatgagggagagaataatgcacaaactgaaacgca 4416
Db 781 atcacatgttgagggccagagcagctgcgtatgagggagagaataatgcacaaactgaaacgca 840
QY 4417 ggtttcaccaacttaggaagcagctgttggagccctgcagctgcagtgtggttagaggg 4476
Db 841 ggtttcaccaacttaggaagcagctgttggagccctgcagctgcagtgtggttagaggg 900
QY 4477 atgggctgaatagggaggttagatttccctgcatacaacagtgcttgggaagcgtgtgga 4536
Db 901 atgggctgaatagggaggttagatttccctgcatacaacagtgcttgggaagcgtgtgga 960
QY 4537 ttccctgaggaagaacaggagccagatggagccacacatgatttgcctcacggctact 4596
Db 961 ttccctgaggaagaacaggagccagatggagccacacatgatttgcctcacggctact 1020
QY 4597 gcagcactttgtaccagaaatctcatgtccacaaaacccatgtaaaacttttcaaccactca 4656
Db 1021 gcagcactttgtaccagaaatctcatgtccacaaaacccatgtaaaacttttcaaccactca 1080
QY 4657 aagctgtttattcgtgtgaagaaataaacttttttttccaccagtcatttgcactctt 4716
Db 1081 aagctgtttattcgtgtgaagaaataaacttttttttccaccagtcatttgcactctt 1140
QY 4717 catatggctgtgtcgacacctccagaaacgtgtgttatacttccagtcagtgtaggagac 4776
Db 1141 catatggctgtgtcgacacctccagaaacgtgtgttatacttccagtcagtgtaggagac 1200
QY 4777 tgaagacttccggttggctgaggaactgaggtttgaccttcgggaagggaagtcccaactca 4836
Db 1201 tgaagacttccggttggctgaggaactgaggtttgaccttcgggaagggaagtcccaactca 1260
QY 4837 tcttatttattatgcctgtgatgtgggtccctgcaggggagacatccagttactcgtgtct 4896
Db 1261 tcttatttattatgcctgtgatgtgggtccctgcaggggagacatccagttactcgtgtct 1320

QY 4897 ttaattgcacacctggggaaactgtgttttatggccttctttggggcactcctcgtgttttggat 4956
Db 1321 ttaattgcacacctggggaaactgtgttttatggccttctttggggcactcctcgtgttttggat 1380
QY 4957 gaagtggagggaalacagaggttaaaagaattgtctccacacctgaagcgggagtcgccgct 5016
Db 1381 gaagtggagggaalacagaggttaaaagaattgtctccacacctgaagcgggagtcgccgct 1440
QY 5017 tcaactttcggaaaatgtgtcagccactggggacagttctgcgcccgggacatggtgttttc 5076
Db 1441 tcaactttcggaaaatgtgtcagccactggggacagttctgcgcccgggacatggtgttttc 1500
QY 5077 tcaaggtctctctaaataataatccctattcttacaataatccttggccctgagtggttttaa 5136
Db 1501 tcaaggtctctctaaataataatccctattcttacaataatccttggccctgagtggttttaa 1560
QY 5137 gcaagaactcctgtgtccccatggtctccacactcaccatcaccctcctgtgtagcaagagt 5196
Db 1561 gcaagaactcctgtgtccccatggtctccacactcaccatcaccctcctgtgtagcaagagt 1620
QY 5197 cctagtcaaggggaggtgcatttttagtattgaaattgcaacttatccatgagataataaaa 5256
Db 1621 cctagtcaaggggaggtgcatttttagtattgaaattgcaacttatccatgagataataaaa 1680
QY 5257 ggagaactgttttatacagtggaggctaaacctaaaaatttcaaaagtctgcgcttttttga 5316
Db 1681 ggagaactgttttatacagtggaggctaaacctaaaaatttcaaaagtctgcgcttttttga 1740
QY 5317 tcttgggcctctctctctgtagaaaccaattggcccttctgtgctcacggcctcgcacctaa 5376
Db 1741 tcttgggcctctctctctgtagaaaccaattggcccttctgtgctcacggcctcgcacctaa 1800
QY 5377 ctggagagtctgtgactcctgcagctcacctgagccacacagactaggtcttcttggctcct 5436
Db 1801 ctggagagtctgtgactcctgcagctcacctgagccacacagactaggtcttcttggctcct 1860
QY 5437 tccgcagcatgctgtctacccccccagaaacccgcagctgtgtgggaagagccatgtaggaggg 5496
Db 1861 tccgcagcatgctgtctacccccccagaaacccgcagctgtgtgggaagagccatgtaggaggg 1920
QY 5497 ctattccacagcatcacacttccactgccttcagctgcagctgcacagctgcacaaatcacttc 5556
Db 1921 ctattccacagcatcacacttccactgccttcagctgcagctgcagctgcacagctgcacaaatcacttc 1980
QY 5557 ctctatcgagccagaaagacttccagctccacaaatgaagtgttctgtcctgaaacactt 5616
Db 1981 ctctatcgagccagaaagacttccagctccacaaatgaagtgttctgtcctgaaacactt 2040
QY 5617 ctgggaagaatcccaacatcgagaaacgggtgtcctgtgagttcccaacatgctcttg 5676
Db 2041 ctgggaagaatcccaacatcgagaaacgggtgtcctgtgagttcccaacatgctcttg 2100
QY 5677 ttcatgggttttcttcgtatggagtggaatgaagtgttttatttttgttcttaactga 5736
Db 2101 ttcatgggttttcttcgtatggagtggaatgaagtgttttatttttgttcttaactga 2160
QY 5737 gaaaaaaggagggcaccacaaaggttgggttcacacagttccacagtttccagggagggcg 5796
Db 2161 gaaaaaaggagggcaccacaaaggttgggttcacacagttcaccagtttccagggagggcg 2220
QY 5797 ttgggggtggggaaggccactccagagcatgaggtcttaaggggacatgagtaaaagcat 5856
Db 2221 ttgggggtggggaaggccactccagagcatgaggtcttaaggggacatgagtaaaagcat 2280
QY 5857 gtcgtgacccagtgaggaaggagagggccagctgcactcctcctgcacggggttccctagctg 5916
Db 2281 gtcgtgacccagtgaggaaggagagggccagctgcactcctcctgcacggggttccctagctg 2340
QY 5917 cagaaagggtcccgccctaggccgagggggaaacacactgtagcagaagaggcctggatgcac 5976
Db 2341 cagaaagggtcccgccctaggccgagggggaaacacactgtagcagaagaggcctggatgcac 2400

QY 5977 acctggcagccaggctctccgcccagacacagtgctccatgtccagccctgcacctgg 6036
|||||
Db 2401 acctggcagccaggctctccgcccagacacagtgctccatgtccagccctgcacctgg 2460
QY 6037 ggtgtgtgattcagctgacagatgacacaaatcctgcaccataatccacagatgggga 6096
|||||
Db 2461 ggtgtgtgattcagctgacagatgacacaaatcctgcaccataatccacagatgggga 2520
QY 6097 agtgagagaaagggaagtgatgtgtaactgtctcaagagatgcttaaacctccataga 6156
|||||
Db 2521 agtgagagaaagggaagtgatgtgtaactgtctcaagagatgcttaaacctccataga 2580
QY 6157 gaggagccggcgaggggcatctgtgtgtcccggtcacacactgcagcaggggaaggtgg 6216
|||||
Db 2581 gaggagccggcgaggggcatctgtgtgtcccggtcacacactgcagcaggggaaggtgg 2640
QY 6217 ctggctggtccctggcatcagtggtttgtttaagctccagagggtcttattgcaattg 6276
|||||
Db 2641 ctggctggtccctggcatcagtggtttgtttaagctccagagggtcttattgcaattg 2700
QY 6277 tcttttctctgccccttgagccagctaaaggccctggagctgtttcttttaggcggatg 6336
|||||
Db 2701 tcttttctctgccccttgagccagctaaaggccctggagctgtttcttttaggcggatg 2760
QY 6337 aactgacatgctctaccatgacaggtctgtggcaggctctgcacaggtctctcagtatcttgaga 6396
|||||
Db 2761 aactgacatgctctaccatgacaggtctgtggcaggctctgcacaggtctctcagtatcttgaga 2820
QY 6397 ggtggcatggaagtgcctattctcaggtacagaaaccttcagagaggataaatagctt 6456
|||||
Db 2821 ggtggcatggaagtgcctattctcaggtacagaaaccttcagagaggataaatagctt 2880
QY 6457 gcctgtagaagcaggaactgaaacccctgtccgcctgaactcccccaactactctgcccac 6516
|||||
Db 2881 gcctgtagaagcaggaactgaaacccctgtccgcctgaactcccccaactactctgcccac 2940
QY 6517 tgtagccccctgcttactgtctgtgacacccctccaccatctctgtataccttaaatatc 6576
|||||
Db 2941 tgtagccccctgcttactgtctgtgacacccctccaccatctctgtataccttaaatatc 3000
QY 6577 aaagagggcgaagagaaagggttttaagataagttatttttttaagggaaccttaatat 6636
|||||
Db 3001 aaagagggcgaagagaaagggttttaagataagttatttttttaagggaaccttaatat 3060
QY 6637 tatttttaagaagtaaccaaattagtgacgtgaaatgcaaaatgcaaaataaaaaa 6692
|||||
Db 3061 tatttttaagaagtaaccaaattagtgacgtgaaatgcaaaataaaaaa 3116
RESULT 9
AAF18333
ID AAF18333 standard; DNA; 3170 BP.
XX
XX
AC AAF18333;
XX
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 352.
XX
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnery;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
OS Homo sapiens.
XX
XX WO20005180-A2.
XX
XX
PD 21-SEP-2000.
XX
XX
PF 08-MAR-2000; 2000WO-US05918.
XX

PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM;
XX
DR WPI; 2000-587514/55.
DR P-PSDB; AAB58457.
XX
PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
PS Claim 1; Page 809-810; 1425pp; English.
XX
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnery; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 3170 BP; 781 A; 758 C; 822 G; 803 T; 6 other;
Query Match 44.3%; Score 3080; DB 21; Length 3170;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 3113; Conservative 5; Mismatches 6; Indels 3; Gaps 3;
QY 3577 gccagagattgtcgggagccacgctgagtgccaagcagggaaggagggcatgcccgg 3636
|||
Db 44 gcacagaatttctcgggagccacgctgagtgccaagcagggaaggagggcatgcccgg 103
QY 3637 gcgtgaccacagtgaggagacagagtgatgtggaacacagagctctcattcagacct 3696
|||||
Db 104 gcgtgaccacagtgaggagacagagtgatgtggaacacagaggtctcattcagacct 163
QY 3697 ttgtgttactgtgaacgtgaatgtggccagtgatcaagagagatctctctgagtgactgc 3756
|||||
Db 164 ttgtgttactgtgaacgtgaatgtggccagtgatcaagagagatctctctgagtgactgc 223
QY 3757 accatggcactggcaccagggcgactattagccaggggcagaccactagacttcagtcgag 3816
|||||
Db 224 accatggcactggcaccagggcgactattagccaggggcagaccactagacttcagtcgag 283
QY 3817 ggaactggtttcccttcctggttcgacttttagtaaatgtgggtggagggttctctttggat 3876
|||||
Db 284 ggaactggtttcccttcctggttcgacttttagtaaatgtgggtggagggttctctttggat 343
QY 3877 ctgtttgagactgttccagaaagaaggtctccttcccagagacacttccatagggcagca 3936
|||||
Db 344 ctgtttgagactgttccagaaagaaggtctccttcccagagacacttccatagggcagca 403
QY 3937 atttggtgattcatttgcagcaaaataactggcttggtaattatttctgccccgcgct 3996
|||||
Db 404 atttggtgattcatttgcagcaaaataactggcttggtaattatttctgccccgcgct 463
QY 3997 gcgtgctaaacaacagatgaggtatgcagctaccactgaagctctgaagatgtcgccattga 4056
|||||
Db 464 gcgtgctaaacaacagatgaggttgcagctaccacwagagctctgaagatgtcgccattga 523

QY 4057 acggacagtggttttcataatgttttotaaggttggtttattgtgtacaggtttccaaagccagcccc 4116
DB 524 acggacagtggttttcataatgtttttaggttggtttattgtgtacaggtttccaaagccagcccc 583
QY 4117 cacagttaggaaatgtgtgagggcacccacacaaact-gcaatgtgttttttaaagtcaagg 4175
DB 584 cacagttaggaaatgtgtgagggcacccacacaaactggcaatgtgttttttaaagtcagaag 643
QY 4176 tgacacatgtatttaaga-ttttttttaaaatctctttgcagttaaatctcaatttttc 4234
DB 644 tgacacatgtatttaagaatgttttttaaaatctctttgcagttaaatctcaatttttc 703
QY 4235 aaacaacgttgatcagggcgaacaaacttatattgttttttagctgagggctcagcagg 4294
DB 704 aaacaacgttcggatcagggcgaacaaacttatattgttttttagctgagggctcagcagg 763
QY 4295 cagattgacggcaggggggcactttttcatcattagggccacagcctggggcctgggactc 4354
DB 764 cagattgacggcaggggggcactttttcatcattagggccacagcctggggcctgggactc 823
QY 4355 tgatcacattgtgagggccagggcgagctgcgtatggaggagaaatgtcaaaactgaacg 4414
DB 824 tgatcacattgtgagggccagggcgagctgcgtatggaggagaaatgtcaaaactgaacg 883
QY 4415 caggtttcacacactctaggaagcagctgtttgaagccctgcagctgagttgattagag 4474
DB 884 caggtttcacacactctaggaagcagctgtttgaagccctgcagctgagttgattagag 943
QY 4475 ggatgggctgaataggcaggttagatttctcgtatcaaacagtgctttgggaagctggtg 4534
DB 944 ggatgggctgaataggcaggttagatttctcgtatcaaacagtgctttgggaagctggtg 1003
QY 4535 gattcctgaggaagaaacagggagccggagatggaggccacacatagtttgcaccggcta 4594
DB 1004 gattcctgaggaagaaacagggagccggagatggaggccacacatagtttgcaccggcta 1063
QY 4595 ctgcagcactttgcaccagaatctcatgtcccaaaacccatgtaaacttcaaacact 4654
DB 1064 ctgcagcactttgcaccagaatctcatgtcccaaaacccatgtaaacttcaaacact 1123
QY 4655 caaagctgtttattcggtgaaagaaataaactttttttctcaccagctcatttgcacctc 4714
DB 1124 caaagctgtttattcggtgaaagaaataaactttttttctcaccagctcatttgcacctc 1183
QY 4715 ttcatatggctgtgtcgccaccctccagaagacgttggtttatacttccagtcagttgggaga 4774
DB 1184 ttcatatggctgtgtcgccaccctccagaagacgttggtttatacttccagtcagttgggaga 1243
QY 4775 actgaagacttcgggttcgaggaactgaggggtgacacttcgggaaggaagtccact 4834
DB 1244 actgaagacttcgggttcgaggaactgaggggtgacacttcgggaaggaagtccact 1303
QY 4835 catcttatttattatgcctgtg-atgtgggttcctgccaggagacatccagtaactcgtgtg 4893
DB 1304 catcttatttattatgcctgtgaaatgtgggttcctgccaggagacatccagtaactcgtgtg 1363
QY 4894 tctttaattgccacctggggaactgtgtttattggcctcttttggggcatcctgggttttg 4953
DB 1364 tctttaattgccacctggggaactgtgtttattggcctcttttggggcatcctgggttttg 1423
QY 4954 gatgaagtggagggaatcacagaggttaaaagaattgtctccacactgaagcgggagtcgcc 5013
DB 1424 gatgaagtggagggaatcacagaggttaaaagaattgtctccacactgaagcgggagtcgcc 1483
QY 5014 gcttcacatttctggaaatggtgcagccactggggacagttctgccccgggcatggttgtt 5073
DB 1484 gcttcacatttctggaaatggtgcagccactggggacagttctgccccgggcatggttgtt 1543
QY 5074 ttcttcaaggtcctctaaatataatccctattcttacaataatccttggccctgatggttt 5133
DB 1544 ttcttcaaggtcctctaaatataatccctattcttacaataatccttggccctgatggttt 1603

QY 5134 taagcaagaactcctcgtgtgtcccatggtctccacacactcaccatcaccctgctgtagcaag 5193
DB 1604 taagcaagaactcctcgtgtgtcccatggtctccacacactcaccatcaccctgctgtagcaag 1663
QY 5194 agtccctagtcaggggaggtgcatttttagtagttaaatgcacttatccatgagataaata 5253
DB 1664 agtccctagtcaggggaggtgcatttttagtagttamattgcacttatccatgagataaata 1723
QY 5254 aaaggagaactgttttttatcagtgggaggtcaacctaaatttcaaaatgtcgcctttttg 5313
DB 1724 aaaggagaactgttttttatcagtgggaggtcaacctaaatttcaaaatgtcgcctttttg 1783
QY 5314 aaatcttgggacctctctctctgtagaaccaatggcccttttggctcaagccctcgacc 5373
DB 1784 aaatcttgggacctctctctctgtagaaccaatggcccttttggctcaagccctcgacc 1843
QY 5374 taactgagagattctgagctcctcagctcactgcagccacagactaggtcttctggct 5433
DB 1844 taactgagagattctgagctcctcagctcactgcagccacagactaggtcttctggct 1903
QY 5434 ccttcgcagcatgctcgtctcaccctccagaccgcgcgtgtgggaagagccaatgtaggg 5493
DB 1904 ccttcgcagcatgctcgtctcaccctccagaccgcgcgtgtgggaagagccaatgtaggg 1963
QY 5494 aggtctattccaggacatacacttccactgccttcagctgcagctcagctcagctgaacaaatcat 5553
DB 1964 aggtctattccaggacatacacttccactgccttcagctgcagctcagctcagctgaacaaatcat 2023
QY 5554 ctctctatcggagccaggaagacttcagctccacaaatgaagtgtctctcctcgaaac 5613
DB 2024 ctctctatcggagccaggaagacttcagctccacaaatgaagtgtctctcctcgaaac 2083
QY 5614 attcttgggaagaaatcccaacatcgagaaacggtgtctcctgtgagttccacaacatgcttc 5673
DB 2084 attcttgggaagaaatcccaacatcgagaaacggtgtctcctgtgagttccacaacatgcttc 2143
QY 5674 ttgtctatgggtttctcctgtatgagtgattaaagtggtttttattttgttcttaac 5733
DB 2144 ttgtctatgggtttctcctgtatgagtgattaaagtggtttttattttgttcttaac 2203
QY 5734 tgagaaaaaaggaggggcacccacaaagttgaggttcacacagctctccacagtttccaggag 5793
DB 2204 tgagaaaaaaggaggggcacccacaaagttgaggttcacacagctctccacagtttccaggag 2263
QY 5794 gcgttttgggggtggggaaggcacctccagagcatgaggctctaagggtctaaggggacatgagtaaaag 5853
DB 2264 gcgttttgggggtggggaaggcacctccagagcatgaggctctaagggtctaaggggacatgagtaaaag 2323
QY 5854 catgtctgtgaccagtgaggaaaggagagggccagctgcactcctgcacggggttccctag 5913
DB 2324 catgtctgtgaccagtgaggaaaggagagggccagctgcactcctgcacggggttccctag 2383
QY 5914 ctgcagaagggtccgcctcctagggccaggggaaacacacatgatagcagaagagggcctgagtg 5973
DB 2384 ctgcagaagggtccgcctcctagggccaggggaaacacacatgatagcagaagagggcctgagtg 2443
QY 5974 cacacctgcacggcggaggtcctccgcccagagacagtgctcctatgtcagccccctgacc 6033
DB 2444 cacacctgcacggcggaggtcctccgcccagagacagtgctcctatgtcagccccctgacc 2503
QY 6034 tggggtgtgtgattccagtgccacagatgccacaatcctgcaccaatattcccacagatggg 6093
DB 2504 tggggtgtgtgattccagtgccacagatgccacaatcctgcaccaatattcccacagatggg 2563
QY 6094 ggaaggtgagagaagggggcaagtgtgttaactgctcaagagatgctttaaactccat 6153
DB 2564 ggaaggtgagagaagggggcaagtgtgtgttaactgctcaagagatgctttaaactccat 2623
QY 6154 agagaggagccggcgaggggcatctgtgttcccgctcacacactgcagcagggaagggg 6213
DB 2624 agagaggagccggcgaggggcatctgtgtgtcccgctcacacactgcagcagggaagggg 2683
QY 6214 tggctggctggctccctcctcagtcagtggtttgttttaagctccagaggggtcttattgcca 6273

Db 2684 tggctggctgcccctggccatcagtggttggtttaagctccagagggtctctatgcca 2743
Qy 6274 ttgtctttctctgccccttgagccagcctaaggccctgagctctgtttctcttaggcgg 6333
Db 2744 ttgtctttctctgccccttgagccagcctaaggccctgagctctgtttctcttaggcgg 2803
Qy 6334 atgaactgacatgctctctaccatgaccaggtctctgggcaaggctctccacagtatcttg 6393
Db 2804 atgaactgacatgctctctaccatgaccaggtctctgggcaaggctctccacagtatcttg 2863
Qy 6394 agaggtgggcatggaagtgcctatttctcaggtacagaaaaccttcagagaggataatag 6453
Db 2864 agaggtgggcatggaagtgcctatttctcaggtacagaaaaccttcagagaggataatag 2923
Qy 6454 ctgtccctgtagagcaggactgaaacctgttccgcgtgactccccagctactctgcc 6513
Db 2924 ctgtccctgtagagcaggactgaaacctgttccgcgtgactccccagctactctgcc 2983
Qy 6514 cactgtagccccctgcttactgtcctgtggcacacccctcaccatcctgtataccttaaat 6573
Db 2984 cactgtagccccctgcttactgtcctgtggcacacccctcaccatcctgtataccttaaat 3043
Qy 6574 atcaagagggaagagagaaagggtttaagataaagtatttttttaaggaaaccttaa 6633
Db 3044 atcaagagggaagagagaaagggtttaagataaagtatttttttaaggaaaccttaa 3103
Qy 6634 tattatttttaagaaagtaaccaaattagtgacgtgaaatgcaaaaaaaataaataat 6693
Db 3104 tattatttttaagaaagtaaccaaattagtgacgtgaaatgcaaaaaaaataaataatn 3163
Qy 6694 gctgact 6700
Db 3164 gctgact 3170

RESULT 10
AAL26437
ID AAL26437 standard; cDNA; 3075 BP.
XX
AC AAL26437;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 18894.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 3478-3479; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
XX activity.
SQ Sequence 3075 BP; 844 A; 653 C; 717 G; 854 T; 7 other;

Query Match 12.0%; Score 831; DB 22; Length 3075;
Best Local Similarity 89.1%; Pred. No. 2.8e-217;
Matches 997; Conservative 0; Mismatches 90; Indels 32; Gaps 8;
Qy 5828 gaggtctaaagggacatgagtaaacatgtctgtgacccagtgaggaaggagagcca 5887
Db 2 gaggtctaaagggacatgagtaaacatgtctgtgacccagtgaggaaggagagcca 61
Qy 5888 gctgcactctgcacaggggttctctagctgcagaagg--tcccgcctagccgagggga 5944
Db 62 gctgcactctgcaggggttctctagctgcagaagggttcccgcgcagggcccgggggg 121
Qy 5945 aaca---cctgatagcagaagggcctggatgcacacctggcacgcgcgaggtctctcgc 6000
Db 122 gacaacacctgaaagcagaagggccctggaaatgaacacctggcacgcgcgagggctc 181
Qy 6001 ccagacacagtgctccatgtcagccctg-----cacctgggggtgtgtgat 6046
Db 182 cccgcgcccaagacacaggtctgcacctgtcaaccccttgcaacctgggtttgtgat 241
Qy 6047 tcaagtg---cacagatgccaaa---tcctgcaccaatatccacagatgggggaag 6098
Db 242 ttcacgttgcaaaaagatgcccaaatcctgcccccccaaaatcccaacagatgggggaag 301
Qy 6099 gtgagaggaaggggcaagtgtgttaactgtcacaagagatgtcttaacctccatagaga 6158
Db 302 gtgagaggaagggcgaagtgtgttaactgtcacaagagatgtcttaacctccatagaga 361
Qy 6159 ggagccgggcgagggggcatctgtgtcccgtcacacactgcagcagggaaagggtgct 6218
Db 362 ggagccgggcgagggcatctgtgtcccgtcacacactgcagcagggaaagggtgct 421
Qy 6219 gctgtgctccctgcatcagtggttgggttaagctccagaggtcttatgcccattgctc 6278
Db 422 ggtgtgctccctgcatcagtggttgggttgaagctccagaggtcttatgcccattgctc 481
Qy 6279 ttttctctgccccttgagccagcctaaggccctggagtgctgtttcttaggcggatgaa 6338
Db 482 ttttctctgccccttgagccagcctaaggccctggagtgctgtttcttaggcggatgaa 541
Qy 6339 ctgacatgctctaccatgaccaggtctctgggcaaggctcctcacagatcctctgagagg 6398
Db 542 ctgacatgctctaccatgaccaggtctctgggcaaggctcctcacagatcctctgagagg 601
Qy 6399 tgggcatggaagtgcctatttctcagatcacagaaaccttcagagagataaataagcttc 6458
Db 602 tgggcatggaagtgcctatttctcagatcacagaaaccttcagagagataaataagcttc 661
Qy 6459 cctgtagaacagcagactgaaacctgtccgctgactccccccagctactctgccactg 6518
Db 662 cctgtagaacagcagactgaaacctgtccgctgactccccccagctactctgccactg 721
Qy 6519 tagccccctgcttactgtcctggcacacccctcaccatcctgtatacccttaataatcaa 6578
Db 722 tagccccctgcttactgtcctggcacacccctcaccatcctgtatacccttaataatcaa 781
Qy 6579 agagggcaagagagaaaggcctttaagataaagttattttttaaggaaaccttaattata 6638
Db 782 agagggcaagagagaaaggcctttaagataaagttattttttaaggaaaccttaattata 841

QY 6639 tttttaagaagtacacaaattagtcgtaaatgc-aaaaaaaaaaaaaaaaaaatgctg 6697
 Db 842 tttttaagaagtacacaaattagtcgtaaatgc-aaaaaaaaaaaaaaaaaaatgctg 901
 QY 6698 actacccttttgaattgcttccagattg-tttttatatgaattcttagacacttg 6756
 Db 902 actacccttttgaattgcttccagattg-tttttatatgaattcttagacacttg 961
 QY 6757 tcatttaagaataagtgcttgcttgcttcagcaagaagacacactggcagctggctttg 6816
 Db 962 tcatttaagaataagtgcttgcttgcttcagcaagaagacacactggcagctggctttg 1021
 QY 6817 gtataggaagtgaaggacacactgggtttctgacaaagtcgagacttaacccttc 6876
 Db 1022 gtataggaagtgaaggacacactgggtttctgacaaagtcgagacttaacccttc 1081
 QY 6877 catctggagagctgggtgcttggcttggctccctgggtagggc 6915
 Db 1082 catct-gagagctgctgcttggcttggctccctgggtagggc 1119
 RESULT 11
 AAK92040
 ID AAK92040 standard; cDNA; 751 BP.
 XX
 AC AAK92040;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human cDNA 5'-end sequence, SEQ ID NO: 500.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayaashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Claim 2; SEQ ID NO 500; 1380pp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is the nucleotide
 CC sequence of the 5'-end of a cDNA provided in the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SQ Sequence 751 BP; 198 A; 187 C; 188 G; 174 T; 4 other;

Query Match 10.2%; Score 708.2; DB 22; Length 751;
 Best Local Similarity 98.7%; Pred. NO. 6.7e-184;
 Matches 743; Conservative 0; Mismatches 7; Indels 3; Gaps 3;
 QY 2699 acgcctctaggagtcacacgcggtggtgattctcactgcaacacacccctttccctggcct 2758
 Db 1 acgcctctaggagtcacacgcggtggtgattctcactgcaacacacccctttccctggcct 60
 QY 2759 ccaatgtgacgctatttgacctggctgtaggattgcagaatagtgtcaactcctgcgaagt 2818
 Db 61 ccaatgtgacgctatttgacctggctgtaggattgcagaatagtgtcaactcctgcgaagt 120
 QY 2819 cctctgtgaggtctgccagctcttgggattctcagagcgcggtatttagagcgggcgagct 2878
 Db 121 cctctgtgaggtctgccagctcttgggattctcagagcgcggtatttagagcgggcgagct 180
 QY 2879 tgtgcaagcggaagagtcgccgaatgtgacaaagacacactcctcactgcactgacactggagc 2938
 Db 181 tgtgcaagcggaagagtcgccgaatgtgacaaagacacactcctcactgcactgacactggagc 240
 QY 2939 gcgttgccctgtgccagtgcaagtcgggatactttcagttcacaagaatggacactcct 2998
 Db 241 gcgttgccctgtgccagtgcaagtcgggatactttcagttcacaagaatggacactcct 300
 QY 2999 gccgagcatgtgaagatgagcttagcttgaataaagaaacactgcactgattgccatttg 3058
 Db 301 gccgagcatgtgaagatgagcttagcttgaataaagaaacactgcactgattgccatttg 360
 QY 3059 gccttgggtgctcaactgtggaaacccctcactcagcttatacctgtggtgacgagccg 3118
 Db 361 gccttgggtgctcaactgtggaaacccctcactcagcttatacctgtggtgacgagccg 420
 QY 3119 cgggaggtgggtcctgctcactcactagggatcgactgattgttaccctgttgcaagaaga 3178
 Db 421 cgggaggtgggtcctgctcactcactagggatcgactgattgttaccctgttgcaagaaga 480
 QY 3179 ataaaaatgacataagaacaaactcatcttcaaaagtggagatttccaaatgtcccatatg 3238
 Db 481 ataaaaatgacataagaacaaactcatcttcaaaagtggagatttccaaatgtcccatatg 540
 QY 3239 ctgaatacccccaaaaatcctcgtctcacaagaatggggcccgagaagctatgaaatgcagt 3298
 Db 541 ctgaatacccccaaaaatcctcgtctcacaagaatggggcccgagaagctatgaaatgcagt 600
 QY 3299 agaatggaagtacacaaacactcctcagatgcaggtgctgactactgcctacaagt 3358
 Db 601 agaatggaagtacacaaacact-ctncagatgacggtgtgtactactcg-ctacaagt 658
 QY 3359 taaggaaatccagaactgaaacgaacggac-tctaccggcctacactgacactgacagga 3417
 Db 659 taaggaaatccagaactgaaacgaacggacttctaccggcctacactgacactgacagga 718
 QY 3418 tcacggcattcttgctatttcccccgacagtat 3450
 Db 719 tcacggcattcttgctatttcccccgacagnat 751
 RESULT 12
 AAK93703
 ID AAK93703 standard; cDNA; 751 BP.
 XX
 AC AAK93703;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human cDNA clone representative sequence, SEQ ID NO: 2163.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1130094-A2.

XX 05-SEP-2001.
PD 07-JUL-2000; 2000EP-0114089.
XX 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX (HELI-) HELIX RES INST.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX Example 11: SEQ ID NO 2163; 1380pp + sequence listing; English.
XX The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used as the
CC representative sequence from a human clone which was used in
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX Sequence 751 BP; 198 A; 187 C; 188 G; 174 T; 4 other;
SQ

Query Match 10.2%; Score 708.2; DB 22; Length 751;
Best Local Similarity 98.7%; Pred. No. 6.7e-184;
Matches 743; Conservative 0; Mismatches 7; Indels 3; Gaps 3;
QY 2699 acgcctctaggagtgccaacgcggtggtgatctcaactgcaaacacacttttccctggcct 2758
Db 1 acgcctctaggagtgccaacgcggtggtgatctcaactgcaaacacacttttccctggcct 60
QY 2759 ccaatgtgacgctatttgacactgagtgatgagtgacgaaatgtgcaactctgcaagt 2818
Db 61 ccaatgtgacgctatttgacactgagtgatgagtgacgaaatgtgcaactctgcaagt 120
QY 2819 cctctgctgaggtctgcagctcttggatctcagagggcggtatcttttagagcgggcagct 2878
Db 121 cctctgctgaggtctgcagctcttggatctcagagggcggtatcttttagagcgggcagct 180
QY 2879 tgtgcaagcggagagtgccgaatgtgcaaaagacacccctccactctgcaactgagcag 2938
Db 181 tgtgcaagcggagagtgccgaatgtgcaaaagacacccctccactctgcaactgagcag 240
QY 2939 gcgttgccctgtgccagtgcaagtcggtgacatttcagttcaacagatggaccactcct 2998
Db 241 gcgttgccctgtgccagtgcaagtcggtgacatttcagttcaacagatggaccactcct 300
QY 2999 gccgagcatgtgaagatggatagccttgaaatgaaacctgcagtggtgccatttg 3058
Db 301 gccgagcatgtgaagatggatagccttgaaatgaaacctgcagtggtgccatttg 360
QY 3059 gccttggtgtcctcaactgtggaacccctatcagcttatcactgtggtgacgcgcgcg 3118
Db 361 gccttggtgtcctcaactgtggaacccctatcagcttatcactgtggtgacgcgcgcg 420
QY 3119 cggaggtgggctcctgctcactcactgagcgcactgactgattgttactctgttcagaaaga 3178
Db 421 cggaggtgggctcctgctcactcactgagcgcactgactgattgttactctgttcagaaaga 480

QY 3179 ataaaaatgacataagcaaaactcatcttcaaaagtggagatttccaaatgtccccatag 3238
Db 481 ataaaaatgacataagcaaaactcatcttcaaaagtggagatttccaaatgtccccatag 540
QY 3239 ctgaatacccccaaaaatcctcgctcacaagaatgggcccagagaagctattgaaatgcag 3298
Db 541 ctgaatacccccaaaaatcctcgctcacaagaatgggcccagagaagctattgaaatgcag 600
QY 3299 agaattggaagtaccataaaacactcctccagatgacggatgtgtactactcctcctacaagt 3358
Db 601 agaattggaagtaccataaaacact-ctncagatgacggatgtgtactactcg-ctacaagt 658
QY 3359 taaggaatccagaactggaacgaacggac-tctaccgccctcacactggactgaccagga 3417
Db 659 taaggaatccagaactggaacgaacggacttctaccgccnctcacactggactgaccagga 718
QY 3418 tcacggcattcttgcatcttcccccggacagat 3450
Db 719 tcacgggattcttgcatcttcccccggacagat 751

RESULT 13
AAH31088
ID AAH31088 standard; cDNA; 869 BP.
XX AAH31088;
AC AAH31088;
XX 27-JUL-2001 (first entry)
XX Human colon cancer cell line Km12L4-A cDNA library derived seq#1022.
DE Human; diagnosis: colon cancer; cancer; malignant; chromosome mapping;
KW detection; colon cancer cell line Km12L4-A; ss.
XX Homo sapiens.
OS WO200018916-A2.
PN 06-APR-2000.
PD 23-SEP-1999; 99WO-US222226.
XX 28-SEP-1998; 98US-01021161.
PR 28-SEP-1998; 98US-01021180.
PR 29-SEP-1998; 98US-0102380.
PR 08-OCT-1998; 98US-0103815.
PR 27-OCT-1998; 98US-0105877.
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI; 2000-293155/25.
XX Polynucleotide library comprising 1079 defined sequences, useful in
PT the form of an array to detect cancer or susceptibility to cancer -
XX Claim 1; Page 479-480; 502pp; English.
PS The present invention describes a library of polynucleotides comprising
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC one of the 1079 sequences; (2) a recombinant host cell containing (1);
CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that
CC specifically binds to (II); (5) a vector comprising (1); and (6) a method
CC of detecting differentially expressed genes correlated with a cancerous
CC state of a mammalian cell comprising detecting a gene product encoded by
CC 65 of the 1079 sequences given in the specification. The polynucleotides

CC are used to monitor patients having (or susceptible) to cancer to detect
CC potentially malignant events at a molecular level before they are
CC detectable at a gross morphological level. The polynucleotides are also
CC useful for monitoring the efficacy of various therapies and preventive
CC interventions. Polynucleotide probes based on the disclosed sequences
CC are useful for chromosome mapping and detection of transcription levels.
CC The 1079 polynucleotide sequences were derived from a human colon cancer
CC cell line Kml12L4-A CDNA library.
XX
SQ Sequence 869 BP; 221 A; 224 C; 235 G; 185 T; 4 other;

Query Match 10.1%; Score 703; DB 21; Length 869;
Best Local Similarity 94.2%; Pred. No. 2e-182;
Matches 826; Conservative 2; Mismatches 34; Indels 15; Gaps 9;
QY 5813 gcacctcagagcatgagctcctaaggagacatgagtaagcatgtctgtgaccacagtga 5872
Dbb 1 gcacctcagagcatgagctcctaaggagacatgagtaagcatgtctgtgaccacagtga 60
QY 5873 ggaaggagagccagctgcactctgcacggggttctctagctgcagaagggtccgcct 5932
Dbb 61 ggaaggagagccagctgcactctgcacggggttctctagctgcagaagggtccgcct 120
QY 5933 aggcggaggggaaacacctgtagcagaaggcctggtggtgcacacctggcagccgaggg 5992
Dbb 121 aggcggaggggaaacacctgtagcagaaggcctggtggtgcacacctggcagccgaggg 180
QY 5993 ctctccgccagacacagtgctccatgctcagccctgcacctggggtgtgtgattcacgt 6052
Dbb 181 ctctccgccagacacagtgctccatgctcagccctgcacctggggtgtgtgattcacgt 240
QY 6053 gcacagatgcacaaatcctgcacaaatattccacagatgggggaa-ggtgagagaagggg 6111
Dbb 241 gcacagatgcacaaatcctgcacaaatattccacagatgggggaa-ggtgagagaagggg 300
QY 6112 gcaagtgtgttaactgtcctaagagatgcttaaacctccatagagagagccggcgca 6171
Dbb 301 gsaagtgtgttaactgtcctaagagatgcttaaacctccatagagagagccggcgca 360
QY 6172 ggggcattgtgttcccgctcacacatgcagcaggggaaagggtggtggtcctcctg 6231
Dbb 361 ggggcattgtgttcccgctcacacatgcagcaggggaaagggtggtggtcctcctg 420
QY 6232 gcacagtggtgttgaagctcacagaggtcttattgcaattgttttctcctcctccc 6291
Dbb 421 gcacagtggtgttgaagctcacagaggtcttattgcaattgttttctcctcctccc 480
QY 6292 cttgagcagcctaaaggcctggagctgtgtttttt-aggcgagatgaactgacatgctcc 6350
Dbb 481 cttgagcagcctaaaggcctggagctgtgtttttt-aggcgagatgaactgacatgctcc 540
QY 6351 taccatgacaaaggcctgtggcgaaggcctcctcacagatccttgagaaggtggcgatg 6406
Dbb 541 taccatgacaaaggcctgtggcgaaggcctcctcacagatccttgagaaggtggcgatg 600
QY 6407 gaagtgcctattctcaggtacagaacaccttcagagaggataaatagc-ttgcctcttag 6465
Dbb 601 gaagtgcctattctcaggtacagaacaccttcagagaggataaatagc-ttgcctcttag 660
QY 6466 aagcaggactgaaacctgtgcccctgactccccagctactctgcccactgtagcccc 6525
Dbb 661 aagcaggactgaaacctgtgcccctgactccccagctactctgcccactgtagcccc 720
QY 6526 ctgacctactgtctggcaccacctcaccatcctgctgataaccttaataatacagaagggc 6585
Dbb 721 ctgacctactgtctggcaccacctcaccatcctgctgataaccttaataatacagaagggc 780
QY 6586 aagagagaaggccttaagataagttatttttttaaggaaaccttaattatttttaa 6645
Dbb 781 aagagagaaggccttaagataagttatttttttaaggaaaccttaattatttttaa 834
QY 6646 gaagtaaccaaatagtgagcgtgaaatgcacaaaaa 6682

Db 835 gaagtaccacaa--tagtggcgtgnaatgccccaaaaa 869
RESULT 14
AAL34960/c
ID AAL34960 standard; cDNA; 537 BP.
XX
AC AAL34960;
XX
DT 08-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 302.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ss.
XX
OS Homo sapiens.
XX
PN WO200155367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01338.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.

PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-451937/48.
 DR P-PSDB; ABB03378.
 XX
 XX Isolated polypeptide for treating, preventing and/or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 XX
 PS Claim 1: SEQ ID NO 302; 781pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pt_sequences.
 XX
 SQ Sequence 537 BP; 137 A; 137 C; 134 G; 123 T; 6 other;

Query Match 7.2%; Score 500.8; DB 22; Length 537;
 Best Local Similarity 97.9%; Pred. No. 5.8e-127;
 Matches 524; Conservative 1; Mismatches 8; Indels 2; Gaps 2;

QY 2633 atgagatcaccaaaacgttaaatatgtttttcagcggtacacgttagtaccgatccta 2692
 |||||
 Db 537 ATGAGATCACCAAAACGTTAAATATGTGTTTTCAGCGTTACNTAGTAACATCGGATCTA 478
 QY 2693 cagtcaacgctctaggagtagtccaaacggtgtgtatctcactgcaaacacctttccc 2752
 |||||
 Db 477 CAGTTCACGCTTCTAGGGAGTCCAAACGCGTGTGATCTCACTCAACACNACNTTTTCCC 418
 QY 2753 tggcctccaatgtgacgctatttgacctggtgacctgtagtagcagaaagtgtcaactcct 2812
 |||||
 Db 417 TGGCCTCCAATGACGCTATTTGACCTGGCTGATAGGATGCAGAAATGTGTCAACTCCT 358

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 18:08:11 ; Search time 6874.72 Seconds
(without alignments)
13648.677 Million cell updates/sec

Title: US-09-840-746-2
Perfect score: 6952
Sequence: 1 gttcgtatgaagaattgcgc.....tactgggaagctatagggtg 6952

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmd:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_hic:*
 - 9: gb_estl:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	698.2	10.0	900	10	BE741056
2	678	9.8	856	10	BG400390
3	650.6	9.4	805	10	BE379009
4	648.8	9.3	745	10	BE876909
5	639.6	9.2	953	10	BG034314
6	631	9.1	745	10	BI223426
7	628.2	9.0	645	10	BI545262
8	625.6	9.0	671	10	BE394218
9	601.4	8.7	630	9	AL596844
10	600	8.6	715	10	BE378473
11	546.8	7.9	566	9	AA121502
12	542.2	7.8	856	10	BF179545
13	541.8	7.8	622	9	AA127811
14	540.4	7.8	740	10	BI090766
15	528	7.6	528	10	BI714757
16	526	7.6	527	9	AW192581
17	520.4	7.5	537	9	AI418494

18	519.6	7.5	912	10	BE515103
19	515.4	7.4	843	10	W28968
20	511.8	7.4	605	9	AW862970
21	506.8	7.3	523	10	BI548253
22	498.6	7.2	522	9	AW294116
23	496	7.1	499	10	BI715315
24	476	6.8	476	10	BF058997
25	475	6.8	475	10	BF058131
26	469.2	6.7	474	10	BE393056
27	467.4	6.7	488	10	BF750474
28	462	6.6	462	9	AI653145
29	457.6	6.6	708	10	W22070
30	457	6.6	457	9	AW275046
31	455	6.5	456	10	BF750527
32	454.8	6.5	456	9	AI829334
33	454	6.5	454	10	BF724576
34	452.8	6.5	467	9	AW207638
35	449.8	6.5	453	9	AI754021
36	449.2	6.5	460	9	AI242664
37	448	6.4	448	9	AI127628
38	445.2	6.4	462	9	AI143494
39	432.4	6.2	434	9	AI624878
40	428	6.2	463	10	BF154764
41	426	6.1	426	9	AA548961
42	424	6.1	424	9	AA600099
43	423.4	6.1	425	9	AA912014
44	423	6.1	429	10	BF724575
45	415.2	6.0	570	10	W52003

ALIGNMENTS

RESULT 1

LOCUS	BE741056	601594040F1 NIH_MGC_9 Homo sapiens	900 bp	linear	EST 15-SEP-2000
DEFINITION	BE741056	mRNA sequence.			
ACCESSION	BE741056	GI:10155048			
VERSION	BE741056.1	EST.			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1. (bases 1 to 900)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-x@mail.nih.gov Tissue Procurement: DCTD/DTP CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Cloned through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov Plate: LLC810 row: P Column: 10 High quality sequence stop: 766.				
FEATURES	Location/Qualifiers				
source	1. .900				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:3948153"				
	/lab_host="DH10B (phage-resistant)"				
	/tissue_type="adenocarcinoma cell line"				
	/notes="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of				

Qy 1756 cccacagagtcacacaaagctg-----taacaaacaaactctctcttggcctccatccttaa 1810
 Db 360 CCCACAGAGTCCACCAAGCTGGTTAAACAAACAAAATTCCTTGGCTCCATCCCTTAA 419
 Qy 1811 cagagtcctccacagagcaaacctctccagccaaagcaccacaaacttagcacaatgtctc 1870
 Db 420 CAGAGTCTCCACAGAGCAAAACCTTCCAGCCACAGCACCACCTTAGCACAATGCTC 479
 Qy 1871 caactttcaaacactacattctgaagacctctcagcctcttatgaccactcctggaccc 1930
 Db 480 CAACCTTTCACAACTACCATTTCTGAAGACCTCTCAGCCCTCTATGACCACCTCTGGCACCC 539
 Qy 1931 tgaacagacagcatctctgtactgacctatagcgtacagactacacactagagaac 1990
 Db 540 TGTCACAGCAGCATCTCTGTGCTACTGGCCCTATAGCCGTACACACTACAGCTGGAAAC 599
 Qy 1991 agctctcgctg-accatcctgaaatac-tagttctctcaaatctcaacagaagtg-aa 2047
 Db 600 AGCTCTCGCTGAACCCATCCTGAAATACTTAGTTCTCTCAAAATCTCAACAGAAGCTGGCCA 659
 Qy 2048 tcagcacagaagaacccagtgattg-tgagtctaccactgattgatccctttga-c 2105
 Db 660 TCACACAGAAAGAACCGAGTGATTGTGATGCTACCACTGGATTTGATCCCTTTGACC 719
 Qy 2106 cagtgtaccc-acatcagcaaaagaaatgaccacaaagcttgcggtitaca-gcagagttac 2163
 Db 720 CAGTGTACCCNACATCAGCAAAAGAAATGACCACAAAGCTTGGCGTTACAGGCAGAGTTA 779
 Qy 2164 agccagcttcaacttccctcctggaaat--ctccttctcccccaaacacacagttgttcca 2221
 Db 780 CAGCCCGTTTCAGGTTCCCTCGGAACATCTCTCTCTCCCCCAACACCAAGTGTCTTCCA 839
 Qy 2222 cggctgaagactggc 2237
 Db 840 CGGTGAAATTTGGTC 855

RESULT 3
 BE379009 805 bp mRNA linear EST 21-JUL-2000
 LOCUS 601236278F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3608586 5',
 DEFINITION mRNA sequence.
 ACCESSION BE379009
 VERSION BE379009.1 GI:9324374
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 805)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-femail.nih.gov
 Tissue Procurement: AFCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LICM261 row: k column: 19
 High quality sequence stop: 793.
 Location/Qualifiers
 1. .805
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3608586"
 /clone_lib="NIH_MGC_44"
 /tissue_type="endometrium, adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"

FEATURES
 source

/note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 214 a 209 c 197 g 185 t
 ORIGIN
 Query Match 9.4%; Score 650.6; DB 10; Length 805;
 Best Local Similarity 97.0%; Pred. No. 2.5e-124;
 Matches 684; Conservative 0; Mismatches 19; Indels 2; Gaps 2;
 Qy 1 gttcgatgaagaattgcccgtttttcaaaaagtggaacagcctcggagatgggaac 60
 Db 95 GTTCGATCAAGAATATGGCGCTTTTCAAAACAAGAGTGGAAACAGCCTCGGAGATGGGAAC 154
 Qy 61 agagaggcgatgggctgtcagaagaatgactgtgcacagccaagagccacacttc 120
 Db 155 AGAGAGGCGGATGGGCTGTGAGAAGATGGACTGTGCACAGCCAAAGAGGCCACCACTTC 214
 Qy 121 ggcttgagcccttcccttctctctgttgagtgagagtgagcgcctctctag 180
 Db 215 GGCTTGGAGCCCGCTCTTCTTCTCTGCTTTGGAGATGGGAGCTGACCAACGCTTCTAG 274
 Qy 181 gaagagaattctctcaggaccagatctctcctggtgcattctacaggaacagcttc 240
 Db 275 GAAGAGAAATTCCTCAGGACCAGATCTCTCTGGCTGCATTTCTACAGGACAGCAGCTTC 334
 Qy 241 ctctcctctctagacttctcctcactcttgaagtacagagaagcttaacactccac 300
 Db 335 CTCTCCCTCTTAGACCTTCTCTCTCTCTGAAAGTACAGAGAAGCTTAACAACTCCAC 394
 Qy 301 tggcctccagagctcctcagtcagtcacaaacaaagacaatgatgtgtacgcgtgtcac 360
 Db 395 TGGCCTCCAGAGCTCCTCAGTCAGTCAAAACAAGACAAATGCATGTTGCCACCGTGTTCAC 454
 Qy 361 tgatggtggccggagaaagcgtgcgactctttgacggtcagctcgggacctgtgacaaagac 420
 Db 455 TGATGGTGGCCCGAGAACGCTGCGATCTTTGACGGTCACTGTGGACCTGTGACGAAGAC 514
 Qy 421 aagaggtctcccaagagactccagaattgccaactctcctcagctcctctcttcacc 480
 Db 515 AGAAGGCTTCCCCAAGGACTCCAGAAATGGCCACGACTTCATCCTCAGTCTCTTTTACC 574
 Qy 481 ctctgcagtgaaatcgagagaagaacagtagagtaactgggaatccaggggagatgaggaatt 540
 Db 575 CTCTGCACTGGAATCGAGAGAAAACAGTAGAGTAACCTGGGAATCCA-GGGATGAGGAATT 633
 Qy 541 cattgaaccatccacagaaaaatgatttggaacttacgctcttgggtggcaaaatgattcc 600
 Db 634 CATTGAACCATCCACAGAAAATGAATGGACTTACGCTTTTGGCGTTGGCAAAATGATTCC 693
 Qy 601 ccaacc-tttggagaacatcagcttgccagcagctcctgagtgcaaaatggaagtcctcat 659
 Db 694 CCACCTTTTGGACAACATCAGCTTGCAGCAGCTCTGAGGTGCAAAATGGAAGTCCCAT 753
 Qy 660 gtctcagactgagactgtgtctagtgactcagtcgacccatgagag 704
 Db 754 GTCTCAGACTGAGACTGTGTCTAGTGTAGTGTGCGCACCCCATGAAAG 798
 RESULT 4
 BE876909 745 bp mRNA linear EST 20-OCT-2000
 LOCUS 601488479F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3890899 5',
 DEFINITION mRNA sequence.
 ACCESSION BE876909
 VERSION BE876909.1 GI:10325672
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 745)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9674 row: n column: 20
High quality sequence stop: 644.
Location/Qualifiers
1. 745
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3890899"
/clone_lib="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."

BASE COUNT 174 a 198 c 217 g 156 t
ORIGIN

Query Match 9.3%; Score 648.8; DB 10; Length 745;
Best Local Similarity 97.4%; Pred. No. 5.9e-124;
Matches 670; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 5838 gggagacatgaaagcatctgtgtaccagtgaggaagggagccagctgacaccc 5897
DB 1 GGGACATAGTAAGGATGCTGTGACCCAGTGAGGAAGGAGGAGCCAGCTGCATCC 60

QY 5898 tgcacgggttcttagctgcagaaaggttcccgctagcgaggggaaacacctgatagc 5957
DB 61 TGCACGGGTTCTTAGCTGCAGAAAGGTTCCCGCTAGCGCGAGGGAAACACCTGATAGC 120

QY 5958 agaagagccctgagatcacacccctggcagcgagggctctccgccagacagagtgtcca 6017
DB 121 AGAAGAGCCCTGGATGCACACCTGGCAGCGGAGGCTCTCCGCCACACACAGTGTCCA 180

QY 6018 tgtcagccctgcacctgggggtgtgtattcacgtgcacagatgcccaatcctgcaacca 6077
DB 181 TGTACGCCCCCTGCACCTGGGGTGTGATTCACGTGCACAGATGCCACATCTGCACCA 240

QY 6078 atatccacagatgggggaaggtgagaggaaggggcaagtgtatgttaactgtctcaagag 6137
DB 241 ATATCCACAGATGGGGGAAGGTGAGAGGAAGGGGCAAGTGATGTGTAAGTGTCAAGAG 300

QY 6138 atgtctaaacctcatagagagagccgaggggcaatctgtgtcccgctcacaca 6197
DB 301 ATGCTTAAACCTCCATAGAGAGAGGCCGGCGCAGGGGCAATCTGTGTCCCGTCACACA 360

QY 6198 ctgcagcaggaaggggtgggtggctccctggcaatcagtggttttaagctcca 6257
DB 361 CTGCAGCAGGAAGGTGGCTGGCTGGCTCCCTGGCATCATGTGTTTGGTTTAAAGCTCCA 420

QY 6258 gagggtcttattgccattgtcttttctctgtcccttgagccagcctaaagggccctgaggt 6317
DB 421 GAGGGCTTATTGCCATTGTCTTTTCTCTCTGCCCCCTTGTAGCCAGCCCTTAAGGCCCTGGAGT 480

QY 6318 ctgtttcttttagcggagtgaaactgacatgctcttaccatgaccagggctctgggcaagggct 6377
DB 481 CTGTTCTTTAGGGGGATGAACCTGACATGCTCTCTACCATGACCAGGCTCTGGGCAAGGCT 540

QY 6378 cctcacagtataccttgagagtggtggcatggaagtgccatttctcaggtacagaaacctt 6437
DB 541 CCTCACAGTATCCTTGAGAGTGGGCATGGAAGTGCCCATTTCTCAGGTACAGAAACTT 600

QY 6438 cagagagataaaatagcttggcctgtagaagcaggaactgaacacctgtgcgcctgactc 6497
DB 601 CAGAGAGGATAAATAGCTTGCTGTAGAAGCAGGACTGAAACCTGTGCCGCTGTATCC 660

QY 6498 cccagctactctgccactgtagcccc 6525
DB 661 CCCAGCTA-TCTGACCACTGTAGCCCC 687

RESULT 5
BG034314 593 bp mRNA linear EST 24-JAN-2001
LOCUS 602302393F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4403831 5',
DEFINITION mRNA sequence.
ACCESSION BG034314.1 GI:12427500
VERSION BG034314
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 953)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10114 row: j column: 24
High quality sequence stop: 614.
Location/Qualifiers
1. 953
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4403831"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: This is a NIH_MGC Library."

BASE COUNT 242 a 237 c 286 g 188 t
ORIGIN

Query Match 9.2%; Score 639.6; DB 10; Length 953;
Best Local Similarity 95.6%; Pred. No. 4.8e-122;
Matches 690; Conservative 0; Mismatches 29; Indels 3; Gaps 3;

QY 5465 ccgcagctgtgggaagcagcatgaggaggctattcccaggcatacacacttccactgcc 5524
DB 1 CCCGACGCTGTGGGAAGAGCCATGTAGGGAGGCTATTTCCAGGCATACACTTCCACTGCC 60

QY 5525 ttaagctgcagctcacagctgacaaatcatctctctctatcgagccagaagaactcagctc 5584
DB 61 TTCAGCTGATGTCACAGCTGACAAATCATCTCTCTATCGGAGCAGGAAGACTTCAGCTC 120

QY 5585 cacaatatgaagtgttct 5644
DB 121 CACAAAATGAAGTGTCT 180

```
QY 5645 cgtgtcctgtgagttcccaaatgcttctgttccatgggtttcttccgtatgagtgga 5704
Db 181 CGGTGCTCTGTGAGTTCCAAACAATGCTCTGTTCATGGGTTTCTTCCGTATGGAGTGA 240
QY 5705 ttaagagtggtttatttgggttcttaactgagaaaaaaggaggccaccacaaggttga 5764
Db 241 TTAAGAGTGTGTTATTTGTTGTTCTTAACCTAGACAAAAAAGGAGGCCACCACAAAGTTGA 300
QY 5765 ggtcacacagttccacagtttccagagagcggtttgggggtgggggaagccacatccagag 5824
Db 301 GGTCCACACAGTCTCCACAGTTTCCAGAGAGCGGTTTGGGGGTGGGGAAGCGACCTCCAGAG 360
QY 5825 catgggctcttaagggggacatgagtaaaacatgtctgtaccctgagaggaagggaggg 5884
Db 361 CATGAGGCTCTAAGGGGACATGAGTAAGCATGTCTGTGACCCAGTGTAGGAAGGGAGAGG 420
QY 5885 ccagctgacctctgcacgggggttccctagctgcagaaggggtccgcctcctagggccgagggga 5944
Db 421 CCAGCTGCACCTCTGCAC -GGGTTCTTAGCTGCAGAAAGGTCGCCCTAGGCCGAGGGGA 479
QY 5945 aacacctgatagcagaagagccgtgatgcacacctggcagccgagggctctccgcccag 6004
Db 480 AACACCTGATAGCAGAAGAGCGCTGGATGCACACCTGGCACGCCGAGGGCTCTCCGCCAG 539
QY 6005 acacagtgctccatgtcagccctgcacctggggtgtgtgattcacgtgcacagatgcc 6064
Db 540 ACACGTGCTCATGTACGCCCCCTGCACCTGGGGGTGTGTGATTTCAGTGCACAGATGCCA 599
QY 6065 caatcctgcaccaata-tcccacagatgggggaagg-tgagaggaaggggcaagtgatgt 6122
Db 600 CAATCTGCACCAATAATTCCTCCAGATGGGGGAAGGTGTGAGGAGGACGGGCCCAATTGAT 659
QY 6123 gtaactgtcaagagatgtctaaacctccatagagagggcgccgaggggcatctgt 6182
Db 660 GTACTGTCTCAGAGATGCTTAAAGCTCCATAGAGAGGAGCCGGGCGCAGGGGCTTTGTGT 719
QY 6183 gt 6184
Db 720 GT 721

RESULT 6
BI223426 LOCUS 745 bp mRNA linear EST 11-JUL-2001
DEFINITION 602943021F1 NIH_MGC_12 Homo sapiens CDNA clone IMAGE:5106125 5',
mRNA sequence.
ACCESSION BI223426
VERSION BI223426.1 GI:14676870
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 745)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11256 row: i column: 06
High quality sequence start: 13
High quality sequence stop: 688.
Location/Qualifiers
1..745
/organism="Homo sapiens"
```

```
FEATURES
source
LOCUS BI545262 745 bp mRNA linear EST 05-SEP-2001
```

```
/db_xref="taxon:9606"
/clone="IMAGE:5106125"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/notes="Organ: cervix; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT 168 a 187 c 188 g 202 t
ORIGIN

Query Match 9.1%; Score 631; DB 10; Length 745;
Best Local Similarity 98.2%; Pred. No. 2.9e-120;
Matches 722; Conservative 0; Mismatches 5; Indels 8; Gaps 8;

QY 4755 ctccagtcagtgaggagaactcgcggttggtgcgaggaactgaggtgacc 4814
Db 12 CTTCAGTCAGTGTGGGAGAACTGAAGACTTCCGGTTGGTCGAGGAACCTGAGGTTGACC 71
QY 4815 ttcgggaaggaagttccactca-tcttatttattatgcctgtgatgtgggtctctgcagg 4873
Db 72 TTCGGGAAGGAAGTTCACACTCACTCTTATTTATATGCTGTGATGTGGTCTCTGCCAG 131
QY 4874 gagacatccagtaactcgg-tgtctttaattgccacctgggaaactgtgtttattgacct 4932
Db 132 GAGACATCCAGTACTCGGATGTCCTTAAATTGCCACCTGGGGAACTGTGTTATTGGCCCT 191
QY 4933 ctttggggcactcctgtgttttggatgaagtggggaatacacagaggtaaaa-gaattgtct 4991
Db 192 CTTTGGGCATCTCTGTTTGGATGAAGTCAAGGGAATACAGAGGTAACGAATTGTCT 251
QY 4992 ccacctgaagcgggagtcctcccttcacatttctggaatggtgcagcactggggaca 5051
Db 252 CCACCTGAAGCGGGAGTCCCGCTTCACATTTCTGAAATGTGTGACGACCTGGGACA 311
QY 5052 gttctgccccggcatggtgttcttcaaggtcctctaaataataacccctattctaca 5111
Db 312 GTTCTGCCCGCGCATGGTGTGTTCTCAAGGTCTCTAAATATATANTCCCTATTCTTACA 371
QY 5112 taatcttggccctgatggtttttaagcaagaactcctgtgtcccatggttccaccactc 5171
Db 372 TAATCTTGGCCCTGTATGTTTAAAGCAAGAACTCTCTGTGCCATGCTCTCCACCACCTC 431
QY 5172 accataacctgtgttagcaagagtcctagtcaggggaggtgcatttta-gtagttaaat 5230
Db 432 ACCATCACCTGCTGTAGCAAGAGTCTCTAGTCAGGGGAGGTGCATTTTACGTAGTTAAAT 491
QY 5231 tgcacttatccatgagataaat-aaaaggagaactgtttttatcagtgagggtcaacct 5289
Db 492 TGCACTTATCCATGAGATAAATAAAAGGAGAACTGTTTTTATCATGAGTGGAGGCTTAACCTA 551
QY 5290 a-aatttcaagtgtgcgctttttgaaatcttgggctctctctctctctgtagaaccaatggc 5348
Db 552 ACAATATCAAGGTGCGCTTTTGTAAATCTTGGGCATCTCTCTCTGTAGAACCAATGGC 611
QY 5349 ccttctgtgctcagcgcctcgcacctaaactggagagttctgagctcctcagctcacctg 5408
Db 612 CCTTTGTGGTCAACGGCTCGCACCTAACCTGAGAGTTCCTGAGCTCTCGAGCTCGAGCTG 671
QY 5409 agccacagactaggtcttctgctcctccgcagatgcctactca-ccccagaacc 5467
Db 672 AG-CCACAGACTAGGCTTCTGGGCTCTCTCCGAGCATGCTGTGTCACCCCGGAGAACCC 730
QY 5468 gcagctgtgggaaga 5482
Db 731 GCAGCTGTGGGAAGA 745
```

```

DEFINITION      603187386F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5259020 5',
ACCESSION       BI545262
VERSION         BI545262.1 GI:15432574
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 645)
AUTHORS         NIH-MGC http://mgc.nci.nih.gov/.
TITLE           National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL         Unpublished (1999)
COMMENT         Contact: Robert Strausberg, Ph.D.
                Email: cgabbs-r@mail.nih.gov
                Tissue procurement: Miklos Palkovits, M.D., Ph.D.
                CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
                Toshlyuki and Piero Carninci (RIKEN)
                CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLNL1452 row: k column: 21
                High quality sequence stop: 645.
FEATURES        1..645
                Location/Qualifiers
                1..645
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5259020"
                /clone_lib="NIH_MGC_95"
                /tissue_type="hippocampus"
                /lab_host="DH10B"
                /note="Organ: brain; Vector: pBluescriptR (modified
                pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcag
                ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTNN-3',
                size-selected for average insert size 2.5 kb and
                normalized to ROT 5. This is a primary library enriched
                for full-length clones and constructed using the
                Cap-trapper method (Carninci, in preparation). Library
                constructed by M. Brownstein (NIH/NHGRI, National
                Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      152 a 149 c 183 g 161 t
ORIGIN
Query Match          9.0%; Score 628.2; DB 10; Length 645;
Best Local Similarity 99.4%; Pred. No. 1.1e-119;
Matches 641; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 3475 agcagaagagactactttaaagtcaggagagagaggactcattgctctgagccag 3534
DB 1 AGCGGAGGAAGACACTACTTTTAAAGTCCAGGAGAGAGAGGGACTCATTTGCTGAGCCAG 60

QY 3535 tcacctggagacctctctcagagagaccgaccaggaggctgcgccagga-tttgtcg99 3593
DB 61 TCACCTGGGACCTCTCTCTCAGAGGACCGCACAGGAGGCTGCCGCCAGGATTTGTGGG 120

QY 3594 agccacgctgagtg9caagcagaagagagagagcagcatcg9ggcgctgaccacagtgag 3653
DB 121 AGCCAGCGTGAGTGGCAGCAGGAGAGGAGCAGGACATGCGGGGCGCTGACCAGTGGAG 180

QY 3654 gagacaggttgatgtggaaccacagcgtcctcattcagcacctttgttactgtgaac 3713
DB 181 GAGACAGGTGGATGTGGAACACAGCGTCTCTATTTCAGCACCTTTGTTACTGTGAAC 240

QY 3714 gtgaatgtgggccaagtatcaagagagctctctctgagtgcactgcaccatggcactggcacc 3773
DB 241 GTGAATGTGGGCGAGTATCAAGAGAGTCTCTCTGAGTGAAGTGCACCATGCGACCTGGCACC 300

QY 3774 agggcgactattagccagggcagaccactagactcttcagtcaggagacctggtttccctt 3833
DB 301 AGGGCGACTATTAGCCAGGTGAGACCACCTAGACTTTTCAGTGCAGGGACCTGGTTTTCCCTT 360

```

```

QY 3834 cgttgcacttagtaaatgggtggaggtttcctcttttgatctgttttgagactgttc 3893
DB 361 CGTTGCACCTTTAGTAATTTGGGTGGGAGGTTTCCCTTTGGATCTGTTTTGAGACTGTT 420

QY 3894 cagaagaagcgtctccttcccgagacacattccatagcagcaaatttgggtgattcattg 3953
DB 421 CAGAAAGAGGCTTCCCTTTCCCGAGACACTTCCATAGCAGCAATTTGGTGATTCATTG 480

QY 3954 cagcaaaataactggctgtgtaattatttctctgcccagcgctgcgtgctaaacaaga 4013
DB 481 CAGCAAAATACTGGCTGTGTAATTTATTTTCTTGCCCCAGCGCTGCTGCTAAACAACA 540

QY 4014 tgaggatgagcgtaccactgaagctcgaagatgctgcattgacgacgacagtgtttcat 4073
DB 541 TGAGGATGAGCGGTACCACTGAAGTCTGAAGATGTGCCATTTGAACGACAGACTGTTTCAT 600

QY 4074 atgtttcaggtgtcttatgctacagtttcccaagccagcccca 4118
DB 601 ATGTTTCTAGTGTGCTTTATGCTACAGTTTCCAAAGCCAGCCCCCA 645

RESULT 8
BE394218
LOCUS      601311526F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3632959 5',
DEFINITION mRNA sequence.
ACCESSION  BE394218
VERSION     BE394218.1 GI:9339583
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 671)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue procurement: ATCC
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM325 row: c column: 08
            High quality sequence stop: 661.
FEATURES    1..671
            Location/Qualifiers
            1..671
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3632959"
            /clone_lib="NIH_MGC_44"
            /tissue_type="endometrium, adenocarcinoma cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: uterus; Vector: pOTB7; Site.1: XhoI; Site.2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5'
            adaptor: GGCACGAG(G). Library constructed by Ling Hong
            in the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT  160 a 179 c 184 g 148 t
ORIGIN
Query Match          9.0%; Score 625.6; DB 10; Length 671;
Best Local Similarity 99.0%; Pred. No. 3.8e-119;
Matches 661; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

QY 5325 ctctctctgtagaaccaaaggccctttgtggtcacggcctgcacctaactggagag 5384

```

```
Db 1 CTCTCTCTCTGTAGAACCAATGGCCCTTTGTGGCTCAGGCCTCGCACCTTAACGTGGAG 60
QY 5385 ttctgaactctcagctcagctcagctcagctcagctcagctcagctcagctcagctcagc 5444
Db 61 TGCTGAGCTCCTCGAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTC 120
QY 5445 atgctcctcaccacccacccacccacccacccacccacccacccacccacccaccc 5504
Db 121 ATGCTCTGCTCACCCTCCAGACCCGACCTGTGGGAAAGCCATGTAGGGAGCTATTCCC 180
QY 5505 aggcataacttccactgccttcagc-tgagctcacagctgacaaatcatctcctctatc 5563
Db 181 AGGCATACACTTCACCTGCCTTCAGCTGAGCTGACGTCACAGCTGACAAATCATCTCCT 240
QY 5564 ggagccagaagacttcagctccacaaatgaagtgtctcctcctgaaacattctggga 5623
Db 241 GGAGCCAGAGAGACTTACGCTCCACAAATGAAGTGTCTGCTCTGAAACATCTCTGGGA 300
QY 5624 agaatcccaacatcgagaaacgggtgctcctgagttcccaaatgcttctgttctatgg 5683
Db 301 AGAATCCACATCGAGAAACGGGTGCTGTGAGTTCCACAAATGCTTCTGTTCATGG 360
QY 5684 gttctccctatggagtgagtaagagtggttttttttttttttttttttttttttttttt 5743
Db 361 GTTTCTTCGCTATGAGTGGATTAAAGAGTGTATTTATTTTGTGTCTTAACGTGAAAAA 420
QY 5744 agagggcaccacaaaggttgaggtcacacagctccacagtttccagggcggtttgggg 5803
Db 421 AGGAGGCCACCCACAAAGTGTGAGGTTCACACAGTCTCCACAGTTTCCAGGAGCG-TT 479
QY 5804 gtggggaagcaccctcagacatgaagctcctaaggagacatgagtaagcatgtctgtg 5863
Db 480 GTGGGGAAGGCACCTCCAGAGCATGAGGCTCTAAGGGGACATGAGTAAGCATGTCTGTG 539
QY 5864 acccagtgaggaagggagagggcagctgcactcctcctgcacggggttctcagtcgaga 5923
Db 540 ACCCAGTGAAGAGGAGAGGCCAGCTGCACCTCTGCACGGGTTCTTAGCTGCAGAGG 599
QY 5924 gtcccgctaggcagggggaacacccctgtagcagaaaggcctgtagtcacacacctggc 5983
Db 600 GTCCCGCTTAGCCGAGGGGGAACACCTGTATGACAG-AGAGGGCTGGATGCACACCTGG 658
QY 5984 acgcccag 5991
Db 659 ACGCCGAG 666

RESULT 9
AL596844 LOCUS 630 bp mRNA linear EST 14-AUG-2001
DEFINITION DKF2p451p091_r1 451 (synonym: hlccl) spinal cord Homo sapiens cDNA
clone DKF2p451P091 5', mRNA sequence.
ACCESSION AL596844
VERSION AL596844
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 630)
AUTHORS Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann,S.).
TITLE EST (Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann
,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Poustka A.J.
Department Lehrach
Max-Planck-Institute for Molecular Genetics
Innestrasse 73, 14195 Berlin, Germany
Tel: +49-30-84131623
Fax: +49-30-84131128
Email: poustka@mpimg-berlin-dahlem.mpg.de
This is the 5' sequence of the clone insert
```

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.
No sl sequence available.
This clone (DKF2p451P091) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers
1. 630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKF2p451P091"
/clone_lib="451 (synonym: hlccl) spinal cord"
/tissue_type="human spinal cord"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT 150 a 201 c 93 g 186 t
ORIGIN

Query Match 8.7%; Score 601.4; DB 9; Length 630;
Best Local Similarity 99.8%; Pred. No. 4e-114;
Matches 602; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1057 ctgtccattacagataaacagttcctcagacatttctctctctctctctctctctctctctctct 1116
Db 28 CTGTCCATTACAGATAACAGTTCATCTCTCAGACATTCTGGAGAGCTCAACTTCTTATATT 87
QY 1117 aaaaatcctcaactcttcacattcagagattctctctctctctctctctctctctctctct 1176
Db 88 AAAATCTCAAACTCTTCCACACTCAGAGTATTCCTCTCTCTCTCTCTCTCTCTCTCTCT 147
QY 1177 agtaaatctcctcctatgacggggaatagctcagcctctctctctctctctctctctctct 1236
Db 148 AGTAACATCTCATCTATGACGGGGAATATGCTCAGGCTTCTACTGAGTCGCCAGTTCTG 207
QY 1237 catactccacaccttcctgctcctacacacccacccaatataatgcccgaacaccttcggtgtt 1296
Db 208 CATACATCCAACTTCCTGCTCTACACACCCACCATTAAATATATGCGGAACACTTCGGTTGT 267
QY 1297 ctggaaactgagctgagttgttagtgactctctctctctctctctctctctctctctctctct 1356
Db 268 CTGGACACTGATGCTGAGTTGTAGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 327
QY 1357 tctctctctcagggcctcctttgcctctgcccctctgttcacaaatcccaacctttattt 1416
Db 328 TCCTCTCTTCAGGGCTCTCTTTCCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 387
QY 1417 tcataaatattaccatcaaacacggcctctgtgcatctactactaaagtctactctgtatgca 1476
Db 388 TCATCAATTTTACCATCAACACGGGCTCTGTGCATCTACTATAAGTCTACCTCTGATGCA 447
QY 1477 tcaacacattggtctctcctcaccatcacccttaaccagatacttaacagacatctacatct 1536
Db 448 TCCACACATGGTCTTCTCTCACCATCACCCTTACCAGTATCTTAAAGACATCTACATCT 507
QY 1537 gcccactttctgtctcacaaaacaccttcccacagtcattctctacccctgtcctgccc 1596
Db 508 GCCCCTCTTCTGTCTCACAACCAACCTTGGCAGAGTCACTCTTACCCCTGTCTCTGCTCTG 567
QY 1597 agggcaaggagagactcctgtgacttcatttcagacatcaacaatgacatcattcatgaca 1656
Db 568 AGGCAAGGAGAGACTCTCTGTGACTTTCATTTCAGACATCAACAATGACATCATTCATGACA 627
QY 1657 atg 1659
Db 628 ATG 630

RESULT 10

BE378473
LOCUS 601236457F1 NIH_MGC_44 715 bp mRNA linear EST 21-JUL-2000
DEFINITION mRNA sequence.
ACCESSION BE378473
VERSION BE378473.1 GI:9323756
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 715)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LCM261 row: b column: 16
High quality sequence stop: 659.
Location/Qualifiers
1..715
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3608367"
/clone_lib="NIH_MGC_44"
/tissue_type="endometrium, adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 162 a 190 c 212 g 151 t
ORIGIN
Query Match 8.6%; Score 600; DB 10; Length 715;
Best Local Similarity 95.9%; Pred. No. 7.8e-114;
Matches 671; Conservative 0; Mismatches 20; Indels 9; Gaps 5;
QY 5827 tgaggctctaaggggacatgagtaaacatgtctgtgacccagtgaaggaggagggcc 5886
Db 1 TGAGGCTCTAAGGGGACATGAGTAAGCATGTCTGTGACCCAGTGAGGAAGGGAGGCC 60
QY 5887 agctgcaactctcagcgggttcttagctcagaagggtcccgccctaggccgaggggaaa 5946
Db 61 AGCTGCACTCTGACGGGGTCTTAGCTGCAGAAAGGGTCCCGCCTAGCGGAGGGAAA 120
QY 5947 cactgtagcagaagagcctgtagtcacacacctggcacccagcagcctcccgccagac 6006
Db 121 CACCTGATACAGAGAGGGCTGTGATGCACACCTGGCACCCGACGGCTCTCCGCCACAC 180
QY 6007 acagtgctccatgtcagccctcacctgggggtgtgtgatttcacgtgcagatggcaca 6066
Db 181 ACAGTGCTCCATGTACGCCCTGCACCTGGGGTGTGTGATTTCAGTGCACAGATGCCACA 240
QY 6067 atctgcaccaatattcccaacagatgggggaagtgtgagagagaaagggcaagtgtgtaa 6126
Db 241 ATCTCGACCAATATCCACAGATGGGGGAAGGTGAGAGAGAGGGGCAAGTGTGTGTA 300
QY 6127 ctgctcaagagatgcttaacctccatagagagagccggcgccagggccatctgtgtgt 6186
Db 301 CTGCTCAAGAGATGCTTAAACCTCCATAGAGAGAGAGCCGGCGCGGAGCATCTGTGTGT 360

QY 6187 cccgtcacacactgcagcagggaagggtggtgctggtgctccctggcatcagtggtttgg 6246
Db 361 CCCGTACACACTGCACAGAGGAGGTGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 420
QY 6247 tttaagctccagagggtcttattgcccattgtctttctctctgcccccttgagccagctaa 6306
Db 421 TTTAAGCTCCAGAGGGTCTTATTGCCATTGTCTTTTCTCTGCCCCCTTGAGCCAGCCTAA 480
QY 6307 ggccttgagctgtttcttttagcgcgatgaactgacatctcctaccatgaccagctc 6366
Db 481 GGCCCTGGAGTCTGTATCTTTTAGCGGATGAACCTGACATGCTCTACATGACCGAGCTC 540
QY 6367 tggg-caaggctcctcagctatctcttgagaggtggcagtggaagtg-ccattttctcag 6424
Db 541 TGGGCCAAGGCTCCTCAGCTATCTTTGAGAGGTGGCATGGAAGTGCCCATTTCTCAG 600
QY 6425 gtacagaaaccttcagagagagataaatagcttgccctgtagaagcagcagactgaaacctt 6484
Db 601 GTACAGAAACCTTCAGAGAGGATAATAGCTTGCCCTGT---AGACGAGGACTGAAACCC--T 655
QY 6485 gtccgctgactccccagctactctctccactgtagcccc 6524
Db 656 TGTGGCTGACTCCCC--AGTAATCTGGCCACTGTAGCCC 693
RESULT 11
AA121502/c 566 bp mRNA linear EST 19-NOV-1996
LOCUS 2k88a11.s1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
DEFINITION IMAGE:489884 3', mRNA sequence.
ACCESSION AA121502
VERSION AA121502.1 GI:1679116
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 566)
REFERENCE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 398.
Location/Qualifiers
1..566
/organism="Homo sapiens"
/db_xref="GDB:3804471"
/db_xref="taxon:9606"
/clone="IMAGE:489884"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
AACTGAGAAGATTCGGCGCGCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 143 a 138 c 142 g 141 t 2 others
ORIGIN

Query Match 7.9%; Score 546.8; DB 9; Length 566;
Best Local Similarity 99.1%; Pred. No. 8.3e-103;
Matches 559; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 6111 gcaagtgatgttaactgtcaagagatcttaaacctccatagagagcgcg 6170
|||
Db 563 GGNCAATGATGTGTAAGTCTCAAGAGATGCTTAAACCTCCATAGAGAGCGCGCGC 504
|||
QY 6171 agggcatctgtgtccctcacactcagcaggaaggggtgctgctgctccct 6230
|||
Db 503 AGGGCATCTGTGTGTCCNGTCACACTGCAGCAGGAAGGGTGGCTGGCTCCCT 444
|||
QY 6231 ggcatcagtggttgggttaagctccagaggtcttattgccattgtctttccctgccc 6290
|||
Db 443 GGCATCAGTGTGTTTAAAGCTCCAGAGGGTCTTATGCCATTGCTTTTCCCTCTGCC 384
|||
QY 6291 ctttagccagcctaagggccctggagctctgtttcttttagcggatgaactgacatgctcc 6350
|||
Db 383 CTTGAGCCCA-CCTAAGGGCCCTGGAGTCTGTTCTTTTAGCGGATGAACCTGACATGCTCC 325
|||
QY 6351 taccatgaccaggtctctggcgaaggtccctcacagtatccttgagaggtggcatgaag 6410
|||
Db 324 TACATGACCAAGGCTCTGGGCAAGGCTCTCACAGTATCTTGGAGAGTGGGATGAAG 265
|||
QY 6411 tgccattcttcagggtacagaaacccctcagagaggataaataagcttgccctgtagaagca 6470
|||
Db 264 TGCCCAATTTCTCAGGTACAGAAACCTTCAGAGAGGATAAATAGCTTGCCCTGTAGAAGCA 205
|||
QY 6471 ggactgaaccccttgctgcgctgactcccccaagctactctgcccactgtagccccctgccc 6530
|||
Db 204 GSACTGAACCCCTTGTGCGGCTGACTCCCCAGCTACTCTGCCACTGTAGCCGCCCTGCC 145
|||
QY 6531 ttactgtccctggcacacccctcaccatcctgtatcaccttaataatacaagaggcaagag 6590
|||
Db 144 TTACTGTCTTGGCACACCCCTCACCATCTGTATACCTTAAATATCAAGAGGCAAGAG 85
|||
QY 6591 agaaaggcctttaaagataagttattttttaaggaaaccttaataatttttaagaagt 6650
|||
Db 84 AGAAAGGCTTTAAAGATAAGTTATTTTAAAGGAACCTTAATATTATTTTAAAGAGT 25
|||
QY 6651 aaccaaattagtcagctgaaatgc 6674
|||
Db 24 ACCAAATTAGTGACGTGAATGC 1
|||

RESULT 12
BF179545
LOCUS 856 bp mRNA linear EST 31-OCT-2000
DEFINITION 601807025F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4037671 5',
mRNA sequence.
ACCESSION BF179545
VERSION BF179545.1 GI:11057687
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 856)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>
Plate: LLAM9315 row: b column: 08
High quality sequence stop: 713.
Location/Qualifiers
source 1. .856
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:4037671"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: Sali;
Site: 2: Noti; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 232 a 223 c 221 g 180 t
ORIGIN

Query Match 7.8%; Score 542.2; DB 10; Length 856;
Best Local Similarity 83.3%; Pred. No. 7.5e-102;
Matches 652; Conservative 0; Mismatches 128; Indels 3; Gaps 3;

QY 2710 gagtccaacggtggtgtagtctcaactgcacaacacccctttccctggcctccaaatgtgacg 2769
|||
Db 1 GAGCCCAAGTAGTACAGTTTTCATCTCCTCACTGAAAAACCACTTTGCCCCCTGGCTGCAATGTGACA 60
|||
QY 2770 ctatttgacctgctgtagatgacagaatgtcacaactcctgcagctcctgactgctgag 2829
|||
Db 61 CTGTTTACCTTGGCAGACAGGATCCAGAAATATGTCAACTCTCTGCAGGTCTCTGCTGTA 120
|||
QY 2830 gtctgacagctcttggtatctcagagcggtatcttttagcgggcagcttgtgcaagcgg 2889
|||
Db 121 GTCTGCCAGCTCTTGGGGTCTCAGAGCGGGGTCTTTCGAGGGGCGACGCTTGTGTAAAGCGG 180
|||
QY 2890 aagagtccccgaatgtgacaagaacacccctccatctgcactgacctggacggtggcctg 2949
|||
Db 181 AAGAGTCCAGAATGTGACAAGAGACCTCCATCTGCAACGACCTGATGTTGCTGCGCTC 240
|||
QY 2950 tgcagtgcaagtcgggatacttccagttcaacaagatgacacactcctccgagcagctg 3009
|||
Db 241 TGTCAGTGCAAGTCTGGCTACTTCCAGTTCAACAAAGATGGATCACTCTTCCCGAGCAATGT 300
|||
QY 3010 gaagatggatataggcttgaaaaatgaaacccctgcagtgagttgcccatttggccttgggtggt 3069
|||
Db 301 GAAGATGGATATAGGCTTGAAAAATGAAACCTGTATGAGTTGCCCATTTGGCCTCGGTGGA 360
|||
QY 3070 ctcaactgtggaacccctccatcagcttaccctggtgtagtcgacgagcggtgaggtggg 3129
|||
Db 361 CTCAACTGTGGAACCCCTATCAGTCTCATCTGTTGATGTCGACGCGCAGGAGGTGGC 420
|||
QY 3130 ctctgtctcatcttagcagtcgactgattgttaccctgttgacagaagaataaaaaatgac 3189
|||
Db 421 CTCTTGTCTATCTTGGCGGTGGCACTCATTTGTACCTTGTTCAGAAAGAGTAAAAATGAC 480
|||
QY 3190 ataagcaaacctctcttcaaaagtggagatttccaaatgttccccaatattgtgatacccc 3249
|||
Db 481 ATAAGTAAACTCATCTTCAAAAGTGGGAGCTTCCAAATGTCCCATACAC-GATGTCCCC 539
|||
QY 3250 aaaaaatcctcgctcacagaatggggccgagaagctattgaaatgcatgagaatggaggt 3309
|||
Db 540 AAGAACTCTCGCTCGCAAGAGTGGGGCCGAGAGCCCATTTGAAATGCGACGAGAGCGGAAGC 599
|||
QY 3310 accaaaaacctctccagatgacggatgtgtactactcgcctacaaagtgaaggaatccca 3369
|||
Db 600 ACCAAAAAATCTCTGAGATGACAGACCCGTGTACTACTCGCCACAAACGTAAAGAAACCT 659
|||
QY 3370 gaacttgaa-cgaacaggactctaccggcctcactggactgactggactgactggcaggtatc 3428
|||
Db 660 GAACTTGAACCGCAAGGACTGTAAACCAAGCTACACTGGACTGCC-GGATCGAGAACTCT 718
|||

/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

```

BASE COUNT      177 a   176 c   185 g   202 t
ORIGIN

Query Match      7.8%; Score 540.4; DB 10; Length 740;
Best Local Similarity 91.2%; Pred. No. 1.8e-101;
Matches 675; Conservative 0; Mismatches 51; Indels 14; Gaps 9;

QY 4742 aaacgtgttatacttccagtcagtggtggagaaactgaacactccggttggtcgagaa 4801
      |||
Db 1 AAAGCTGTTATATCTCCAGTCAGTCGTGGGAGAACTGAAGACTTCGGTGGTCGAGGAA 60
      |||

QY 4802 ctgaggggtgacctctcgaggaaggaagtcccaactcattattattatgcctgtgatg 4861
      |||
Db 61 CTGAGGGTTGACCTTCGGGAAGGAAGTTCACACTCATCTTATTATTATGCTGTGATGTG 120
      |||

QY 4862 ggtcctccagggagacatccagtactcgtgtgcttcttaattgccacctgggaactgtg 4921
      |||
Db 121 GGTCTCCAGGAGAGACATCCAGTACTCGGTGCTTTAAATTGGCCACCTGGGGAACTGTG 180
      |||

QY 4922 ttattggcctcttttggggcatcctcgtgttttggatgaagtgaagggaatacagaggttaa 4981
      |||
Db 181 TTATTGGCCCTCTCTTGGGGCATCTCTGTTTGGATGAAGTGAGGGGAATACAGAGGTAAA 240
      |||

QY 4982 agaattgtctccacctgaagcggggagtcctccgcttcacatttctggaattggtgcagcc 5041
      |||
Db 241 AGAATTGCTCCACCCCTGAAGCGGGAGTCCCGCTTCACATTTCTGGAATGGTGCAGCC 300
      |||

QY 5042 actggggacagtctgc-cccgggcatggt-tgtttctcaaggctcctaaataatc 5099
      |||
Db 301 ACTGGGGACAGTCTGCACCCGGGCATGGTATGTTCTCAAGTCTCTACGCTATAATC 360
      |||

QY 5100 cctattctcataaactcttg-gcctgtatggttttaagcaagaactcctgtgtcccatg 5158
      |||
Db 361 CTTATTCTTACATAAATCTTGAGCCCTGATGGTTTAAAGCAAGAACTCTGTGTCCCATG 420
      |||

QY 5159 gtctccaccactcaccatcaccctgtgtgtagcaagagtcctagtcaggggaggtgcatt- 5217
      |||
Db 421 GTCTCCAGCACTCACCATCACCTCGTGTGAGCAAGAGTCTCTAGTCAGGGGAGGTGCTATTA 480
      |||

QY 5218 ttagttagttaaattgcaattatcc-atgagataaaataaaggagaaactgttt--ttatca 5274
      |||
Db 481 TAAGTAGTTACATGGCACTTATCAATGAGATAAATAACAAGGAGAACTGTTACTATCCA 540
      |||

QY 5275 gtgagggttaacctaaattcaaaagtgcgcctttttgaaattcttgggc---ctctctc 5331
      |||
Db 541 GTGGAGGCTAACTTAAATTTCAAGTGTGCGCTTTTAAAGAACTTGGGGCCATCTCTCTC 600
      |||

QY 5332 tctgtagaaccaatggcccttttgg---ctcacggcctcgccactaac-tggagagttc 5387
      |||
Db 601 TGTAGAAACCAATGGGCCCCTTTGTGGGATCACGGGCTCGACACACTAACTTGGAGAGTTC 660
      |||

QY 5388 tgagctcctcagctcacttgagcccaacagactaggcttcttggctccttcgcagcatg 5447
      |||
Db 661 TGAGCTCCTGTAGCTACCTTGACACACACAGACTATGCTCTTGCTCCATCCGGAAGCATG 720
      |||

QY 5448 cctgctcaccctccagaaccc 5467
      |||
Db 721 CTTGCTAAACCCAGAACCGC 740
      |||

```

RESULT 15

BI714757/c 528 bp mRNA linear EST 19-SEP-2001
 LOCUS BI714757
 DEFINITION Icfte03.x1 HR85 islet Homo sapiens cDNA 3', mRNA sequence.
 ACCESSION BI714757
 VERSION BI714757.1 GI:15690452
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 528)
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
 Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
 M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.,
 Jackson, Y., and Bowers, Y.

TITLE

Endocrine Pancreas Consortium
 Unpublished (2000)

JOURNAL

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138

Tel: 617-495-1812
 Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center for information on
 obtaining a clone please contact: Dr. Hiroshi Inoue
 (hinoue@im.wustl.edu)

Seq primer: -400P from Gibco
 High quality sequence stop: 446.

FEATURES

Location/Qualifiers
 1..528
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HR85 islet"
 /tissue_type="Purified pancreatic islet"
 /lab_host="DH10B"
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
 NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
 Size-selected on agarose gel. Average insert size ~1kb. 5'
 XhoI site was destroyed after directional cloning.
 Amplified once. Contact information: Hiroshi Inoue, MD,
 Metabolism Div. (Alan Permutt Lab), Washington University
 School of Medicine, Box 8127, 660 South Euclid Ave., St.
 Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
 314-362-1916, Fax: 314-747-2692."

BASE COUNT

134 a 130 c 137 g 127 t

ORIGIN

Query Match 7.6%; Score 528; DB 10; Length 528;
 Best Local Similarity 100.0%; Pred. No. 6.6e-99;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 6148 ctccatagagaggagcggcgaggggcatctgtgtcccgtcacacactgcagcagg 6207
      |||
Db 528 CTCATAGAGAGGAGCGGGCGAGGGGCATCTGTGTCTCCGTCACACACTGGCAGCAGG 469
      |||

QY 6208 gaaggggtgctggtcgtccctgcatcagtggtttggttaagctccagaggtctta 6267
      |||
Db 468 GAAGGGTGGCTGGCTCGCTCCCTGGCATCAGTGGTTTGGTTTAAGCTCCAGAGGCTCTTA 409
      |||

QY 6268 ttgcatgtctttctctgccccttgagccagccctaaaggcccttgagctgtttcttt 6327
      |||
Db 408 TTGCCATTGTCTTTTCTCTGCCCCCTTGAGCCAGCCCTAAGGCCCTGGAGTCTGTTCTTT 349
      |||

QY 6328 agcgagatgaactgacatgctcctaccatgacacaggtcctgggcaaggctcctcacagta 6387
      |||
Db 348 AGCGGATGAACATGACATGCTCTTACCATGACCAGGCTCTGGGCAAGGCTCTCTCACAGTA 289
      |||

QY 6388 tccttgagagtggtgcatgaagtgcacattctcagttacagaaaccttcagagagat 6447
      |||
Db 288 TCCTTGAGAGTGGGATGGGAAGTGCCCATTTCTCAGGTACAGAAACCTTCAGAGAGGAT 229
      |||

QY 6448 aatagcttgcctgtagaagcaggaactgaaacctctgtccgctgactccccagctac 6507
      |||
Db 228 AATAGCTTGGCCCTGTAGAGCAGGACTGAACCCCTGTGTCGCCCTGACTCCCCCAGCTAC 169
      |||

```

Search completed: July 25, 2002, 22:06:58
Job time: 14327 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 20:03:31 ; Search time 192.67 Seconds
(without alignments)
8863.051 Million cell updates/sec

Title: US-09-840-746-2
Perfect score: 6952
Sequence: 1 gttcgaatgaagattgcgc.....tactgggaagctatagggtg 6952

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCRUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5098.4	73.3	6407	2	US-08-616-844-7
2	5098.4	73.3	6407	2	US-08-599-654-7
3	5098.4	73.3	6407	3	US-08-944-868A-7
4	5098.4	73.3	6407	3	US-08-944-423A-7
5	5098.4	73.3	6407	3	US-08-944-496-7
6	470.4	6.8	584	4	US-09-328-111-83
7	126.4	1.8	128	2	US-08-480-994-7
8	126.4	1.8	128	2	US-08-485-573-7
9	126.4	1.8	128	3	US-08-925-743-7
10	126.4	1.8	128	4	US-08-925-767-7
11	89.8	1.3	7218	1	US-08-232-463-14
12	56.2	0.8	3489	2	US-08-728-323A-1
13	56.2	0.8	32207	2	US-08-770-379-20
14	56.2	0.8	32207	4	US-08-757-669A-20
15	56.2	0.8	32207	4	US-09-230-371A-20
16	52.8	0.8	2518	4	US-09-433-699-3
17	49.4	0.7	1236	2	US-08-741-134-5
18	48.8	0.7	289	4	US-09-007-005-17
19	48.8	0.7	289	4	US-09-244-796-17
20	48.8	0.7	696	4	US-09-461-697-193
21	48.8	0.7	696	4	US-09-461-697-191
22	48.8	0.7	717	4	US-09-461-697-189
23	48.8	0.7	774	4	US-09-461-697-187
24	48.8	0.7	819	4	US-09-461-697-185
25	48.8	0.7	1669	4	US-09-461-697-184
26	48.6	0.7	1931	2	US-09-130-114-2
27	48.2	0.7	966	2	US-08-766-738-2

c 28	48.2	0.7	3211	2	US-08-574-959A-8	Sequence 8, Appli
c 29	48.2	0.7	3211	4	US-09-357-014-8	Sequence 8, Appli
c 30	48.2	0.7	3901	2	US-08-574-959A-6	Sequence 6, Appli
c 31	48.2	0.7	3901	4	US-09-357-014-6	Sequence 6, Appli
c 32	46.6	0.7	1052	1	US-08-466-603-1	Sequence 1, Appli
c 33	46.6	0.7	1052	1	US-08-314-503A-1	Sequence 1, Appli
c 34	46.6	0.7	1052	2	US-08-468-066-1	Sequence 1, Appli
c 35	46.6	0.7	1052	2	US-08-466-717-1	Sequence 1, Appli
c 36	46.6	0.7	1052	3	US-08-466-743-1	Sequence 1, Appli
c 37	46.6	0.7	1052	5	PCT-US95-12414-1	Sequence 1, Appli
c 38	46	0.7	18596	4	US-09-318-448-11	Sequence 11, Appli
c 39	44.8	0.6	9636	1	US-08-323-170B-1	Sequence 1, Appli
c 40	44.8	0.6	9636	4	US-08-954-441-1	Sequence 1, Appli
c 41	44.2	0.6	5163	3	US-08-700-651-1	Sequence 1, Appli
c 42	44.2	0.6	5163	3	US-08-928-361B-4	Sequence 4, Appli
c 43	44.2	0.6	5318	3	US-08-700-651-2	Sequence 2, Appli
c 44	44.2	0.6	5318	3	US-08-928-361B-3	Sequence 3, Appli
c 45	43.6	0.6	43795	3	US-08-742-185-101	Sequence 101, App

ALIGNMENTS

RESULT 1
US-08-616-844-7
; Sequence 7, Application US/08616844
; Patent No. 5849578
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/616,844
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6407 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-616-844-7

Qy	1800	tcaatcttaacagagctctccacagagcaaaccttccagccacacagcaccacttagc	1859
Db	2766	TCCATCCTTTAACAGAGCTCTCCACAGAGCAAAACCTCTCCAGCCACCAAGCACCACTTAGC	2825
Qy	1860	acaaatgtctccaaacttccaaactaccattctgaagacctctcagctctctatgaccac	1919
Db	2826	ACAAATGCTCCAACTTTCCAACTTACCANTTCTGAAGACCTCTCAGCCTCTTTATGACCAC	2885
Qy	1920	tcttggcacacctgtcaagcacagcatctctgttgaactggccctatagcgttacagactac	1979
Db	2886	TCCTGGCACCCCTGTCAAGCACAGCATCTCTGTCTACTGGCCCTATAGCCGTACAGACTAC	2945
Qy	1980	agctggaaaaacagctctcgtgaccatcctgaaatactagttctctcaatctcaacaga	2039
Db	2946	AGCTGGAAAAACAGCTCTCGCTCAGCCATCTCGTGAATACTAGTTGCTCAAAATCTCAACAGA	3005
Qy	2040	aggtggcatcagcacagaaggaacgagtgatgtgatgctaccactggatgatccc	2099
Db	3006	AGGTGGCATCAGCACAGAAAGNACCGAGTGATGTGGATGCTACCACTGGATTTGATCCC	3065
Qy	2100	tttgaccagtgtaaccacatcagcaaaagaaatgaccacaaagctctggcgtttacagcaga	2159
Db	3066	TTTGAACAGTGTACCACATCAGCAAAAGAAATAGCCACAAGCTTGGCGTTACAGCAGA	3125
Qy	2160	gtacagccagcttcaagtcacctgggaacatctcctctcccaaacacacagttgttc	2219
Db	3126	GTACAGCCAGCTTCACGTTCCCTCGGAACATCTCCTTTCTCCCAAAACACAGTTGTTTC	3185
Qy	2220	cacggtggaagacttggctcccaaatcttgccaccttgtctgcagagcagcacacagtc	2279
Db	3186	CAGCGCTGAAGACTTGGCTCCCAATCTGCCACCTTTGCTGTTGAGAGCAGCACACAGTC	3245
Qy	2280	accacaacactgtctctcttcagctcagtcacacagctgtgtgtaacctgtcttca	2339
Db	3246	ACCAACAACACTGTCTCTTTCAGCTCTCAGTCAACAGCTGTGCTGTGAACCCCTTGCTTCA	3305
Qy	2340	caatggcgaatgcgtgcagacacacacacagcgtgtgtctacacatgcagggtcccgccttc	2399
Db	3306	CAATGGCGAATGGCTCGCAGACAACACACGCGTGGCTTACCACATGCAGTGCAGCGCCCTTC	3365
Qy	2400	ctggcaagggaatgattgcagtgtagtgatgaatgagtgctgtcgaaacctctgccatc	2459
Db	3366	CTGGCAAGGGGATGATTGCAGTGTGGATGTGAATGAGTGCCTGTGCAACCCCTGCCCATC	3425
Qy	2460	cacagccacgtgcaacaa tactcagggatccctttatctgcaaatgcccggttgggtacca	2519
Db	3426	CACAGCCACGTGCAACAATACTCAGGGATCCTTTTATCTGCAAAATGCCCGTGGGTACCA	3485
Qy	2520	gttggaaaaagggatagcaatttggttagaaacctgtgcagagatttaaataaagag	2579
Db	3486	GTTGGAAAAAGGGATATGCAATTTGGTTAGAACCTTCGTGACAGATTTTAAATTTAAAGAG	3545
Qy	2580	aactttcttaatacaactgtgaaaaacattcagacctcaagaagttaaaatagat	2639
Db	3546	AACCTTTCTTAATACAACTGTGGAAAAACATTACAGCCTACAGAAGTTGAAAAATGAGAT	3605
Qy	2640	cacaaaaacgttaaatatgttttttcagctgttaactagttacatccgatctacagttca	2699
Db	3606	CACCAAAACGTTAAATATGTTTTTTCAGCGTTCCTAGTTTACATCCGATCTACAGTTCA	3665
Qy	2700	cgctcttagggagtcacaacggtgtgatctcactgcacaaacaccttttcccttggcctc	2759
Db	3666	CGCCTCTAGGGAGTCCAAACGCGGTGGTGATCTACTGTGCAAAACAACCTTTTCCCTGGCCTC	3725
Qy	2760	caatgtcagctatttgacctggctgattaggatgcagaaatgtgtcaactcctgcaagtc	2819
Db	3726	CAATGTGACGCTAATTTGACCTGGCTGTATGAGGATGCAGAAATGTGTCAACTCCTTCAAGTC	3785
Qy	2820	ctctgctgaggtctgcagctcttgggatcttcagagcgaggtatcttagagcgggcagctt	2879
Db	3786	CTCTGCTGAGGTCTGCCAGCTCTTGGGATCTCAGAGGCGGATCTTTAGAGCGGGCAGCTT	3845
Qy	2880	gtgcaagcggaagagtcgccgaatgtgacaagaagacaccttccatctgcactgacctggacg	2939

Db	3846	GTGCAAGCGGAAGAGTCCCGAAATGTGCACAAAGACACTTCATCTGCATGCACCTGGACGG	
Qy	2940	Cgttgcccttgcagtcgcaatcgagataacttcaagtccaagaatggaccactcctg	
Db	3906	CGTTGCCCTGTGCCAGTGCGAAGTCGGGATCTTTTCATTGTTCAACAAGATGGACCACCTCCTG	
Qy	3000	cggagcatgtgaagatggatataggcttgaaaatgaaaacctgatgttgcacatttg	
Db	3966	CCGAGCATGTGAAGATGGATATAGGCTTGAAAATGAACCTGCATGAGTTGCCCATTTTG	
Qy	3060	ccttgggtctcaactgttgaaaaccccctatacagctttacatgtgttgatgcagcgc	
Db	4026	CCTTGGTGGTCTCAACTGTGGAAACCCCTATACAGCTTATCACGTGTGTGATCGCAGCCGC	
Qy	3120	gggaggiggcctcctcctcatcttaggcacatgcgactgatgttacctgtttcacaaaaaa	
Db	4086	GGGAGGTGGGCTCCTGCTCATCTTAGCATCGGACGTGATGTACTGTTGTCAGANAAGNA	
Qy	3180	taaaaatgcataaagcaaacctcatcttcaaaagtggagatttccc aaatgcccatatgc	
Db	4146	TAAAAATGACATAAGCAAACCTCATCTTCAAAAGTGGAGATTTCCA AATGTCCCCATATGC	
Qy	3240	tgaataccccaaaaatcctcgtctcaaaagaatggggccgagaagctattgaaatgcataga	
Db	4206	TGAATACCCC AAAAATCCTCGCTCA CAAGAATGGGCCGAGAAGCTATTTGA AATGCAATGA	
Qy	3300	gaatggaaagtaccaaaaaacctcctccagatgcagcagatgtgtactactcgctctacaagtgt	
Db	4266	GANTGGAAAGTACCAAAAACCTCCTCCAGATGACGGATGTGTACTACTCGCTTACNA GTGT	
Qy	3360	aaggaatccagaacttgaa cga aacggactctaccggcgctacactggactggcaggatc	
Db	4326	AAGGAATCCAGAACTTGAACGA AAACGGACTCTACCCGGCGCTACACTGGACTGCCAGGATC	
Qy	3420	acggcatcttgcatttccccggacagataaacccgtctttcatcagtgatgaagcag	
Db	4386	ACGGCATCTTTGCAATTTCCCGGACAGATTAACCCGCTTTTCATCAGTGATGA AAGCAG	
Qy	3480	aagaagagactacttttaagtccagagagagagggaactcatgtctlgagccagtcaac	
Db	4446	AAGAAGAGACTACTTTTAAGTCCAGNAGAGAGGGGACTCATTTGCTCTGAGCCACTCACC	
Qy	3540	tgggaacctctgctoa gaggaccgcacagaggctgcgccaggatttgtcggaggccac	
Db	4506	TGGGACCTCTGCTCAGAGGACCGCACAGAGGGCTGCGCCAGGATTTGTCGGGAGCCAC	
Qy	3600	gctgagtgcaagcaggaagagggacagcatacgggggcgtagacacagtgagagagaaca	
Db	4566	GCTGAGTGGCAAGCAGGAAGAGGGACAGGCATGCGGGGGCGTGACCACAGTGGAGGAGACA	
Qy	3660	ggtgagatggaaccacagcgtcctcatlcagcaccttggttgactgtgaagctgaat	
Db	4626	GGTGATGTGGAACCAACAGGCTCTCATTCAGCACCTTTGTTGTACTGTGAACGTGAAT	
Qy	3720	gtggggcagLatcaagagagctctctgtagtgactgcacatggcactggcaccaggcg	
Db	4686	GTGGGCCAGTATCAAGAGAGTCTCTGAGTGACTGCACCATGGCAGTGGCACCAGGGCG	
Qy	3780	actattagccaggccagacaactagacttcagtgaa ggaacctggtttccctcttgttg	
Db	4746	ACTATTAGCAGGGCAGCACCATFAGACTTTCAGTGCAAGGACCTGGTTTTTCCTCTCGTTTG	
Qy	3840	cactttagtaaatgggtgggaaggtttccttttgatctgttttgagactgttccaga aa	
Db	4806	CACTTTAGTAATGGGTGGAGGTTTCCTTTTGATCTGTTTTTGAGACTGTTCACAGAA	
Qy	3900	gaaggctctcttcccgagacacttccatlaggcagcaaatgttggtattcatcttcagcaa	
Db	4866	GAAGGCTTCCTTTCCCGAGACACTTCATAGGCAGCNA TTTGCTGATTCATTTGCASCNA	
Qy	3960	aatactggctgtttaattttcttcggccagcgcctcgtgcttaaacacacagatgagga	

Db 4926 AATACGTGGTGTAAATATTTTCCGCCAGCRCCTGGCTGCTAAACAACAGATGAGGA 4985
 QY 4020 tgaagctaccactgaagctctgaagatgtcgccattgaacggaacagtggttttccatattgttt 4079
 Db 4986 TGASCGTACCACCTGAAGTCTGAAGATGTGCCATTTGAACGGACAGTGTGTTTCATATGTTT 5045
 QY 4080 ctagggtgtcttatgtctacagtttccaagccagcccccacagtgaggaatgtgtgagggc 4139
 Db 5046 CTAGGTGTCTTATGCTACAGTTTCCAGCCASCSCCCACAGTGAAGAAATGTGTGAGGC 5105
 QY 4140 accgcacacaactgcgaatgtgttttttaagtcgaagtgacacatgtatttaagatttttt 4199
 Db 5106 ACCGCACACAACCTGCAATGTGTTTAAAGTCAAGGTGACACATGTATTAAAGATTTT 5165
 QY 4200 tttaaaatctcttcagtttaaatctcaacttttttaaaagcctggatcaggggcaaaac 4259
 Db 5166 TTTAAATCTCYTGGCAGTTAAATCTCACTTCTTAAACAAGCCTGGATCAGGGCAAAAC 5225
 QY 4260 aacttatattgtgttttagctggaggtcagcagggcagattgaggcagggggcacttt 4319
 Db 5226 AACTTATATVGTGTTTACGTGGAGGCTCAGCAGGCAGATTGCGAGCGGGGCGACTTT 5285
 QY 4320 tcatcatagggccacagcctggggcctgggactctgataccactgtgagggccagagg 4379
 Db 5286 TCATCCATGAGGGCCAGCCTGGGGCTGGGACTCTGATCACCATTGTGGAGGCCAGAGG 5345
 QY 4380 cagctcgatgagagagaaatgcaaaactgaacgaggtttcaccactctaggaaagca 4439
 Db 5346 CACTGCGTATGAGGAGAAATGTCAAACTGAAACGGAGGTTTCCACCTCTAGGAAGCA 5405
 QY 4440 gcttgttgagccctcagctgagtgatgtggttagagggatgggctgaatggcagggttaga 4499
 Db 5406 GCTTGTGACCCCTGCASCCTGCGATGTGTTACAGGGATGGGCTGAATAGSCAGGTTAGA 5465
 QY 4500 ttccctgcatcaacagtgcttgggaagctgtgtgattctctgaggaagaacagggagcc 4559
 Db 5466 TTTTCCCTGCAACAGTGTCTTGGGAASCTGTGTGATTCTCTGAGGAAGAACAGGAGCC 5525
 QY 4560 gagatgagccacacatgagttgtctcaccggtactgagcacttgtaccagaaatct 4619
 Db 5526 GAGATGGAGCCACACATGAATTTGCTCACCAGGCTACTGACGACATTTGTACCCAGAACTCT 5585
 QY 4620 catgtccacuaacccactgtaaactttcaaccactcaaaagctgtttattcggctgaagaa 4679
 Db 5586 CATGTCCACAACCCCTGTAATCTTCAACCACTCAAAAGSTGTTTATTCGGCTGAAGAA 5645
 QY 4680 ataactttttctcaccagtcattgtacactttcatatgctgtatgcacactcc 4739
 Db 5646 ATAACTTTTCTTCTCACCAGTCATTTGTACCTCTTCATATGGTATGTCGCCACCTCC 5705
 QY 4740 agaaacgtggttatacttccagtcagtggtgggagaactgaagacttccggttgctgag 4799
 Db 5706 AGAAACGTGTTTACTKCCAGTCAGTGTGGGAGAACTGAAGACTTCCGTTTGGTCGAGG 5765
 QY 4800 aactgaggggtgaccttcgggaagtgatgtccactcatcttattattatgcctgtgatg 4859
 Db 5766 AACTGAGGGTGTACCTTCGGGAAGAGTTCACCTCATCTTATTTATATGCTGTGATG 5825
 QY 4860 tgggtcctgcaggagacatccagttactcgtgtgttttaattgccacctggggaactgt 4919
 Db 5826 TGGGTCTGCCAGGGAGACATCCAGTACTCGGTGTCTKTAATTTGCCACCTTGGGAACTGT 5885
 QY 4920 gtttattggcctcttttggggcatcctgggttttggatgaagtgaagggaatacagaggta 4979
 Db 5886 GTTTATTGGCTTCTTTGGGCACTCTGGKTCGGATGAAGTGAAGGGAATACAGAGGTA 5945
 QY 4980 aagaattgtctccacctgaagcggggagtcoccttcaactttcttggaaaatggtgcag 5039
 Db 5946 AAAAATTTGCTCCACCTTGAAGCGGGAGTCCCGCTTTCACATTTCTTGGAAATGTTGCAG 6005
 QY 5040 ccactgggagacagttctgccccgggcatgtgtgttcttcaaggtcctctaaataaac 5099
 Db 6006 CCACTGGGAGACAGTTCTTGCCCGGGCATGTTGTTTCTTCAAGGTCTCTTAAATAAATC 6065

QY 5100 cctattcttaataactcttggccctgatggttttaagcaagaactcctctgtcccatgg 5159
 Db 6066 CCTATTCTTACATAATCTTTGGCCCTGATGTTTAAAGCAAGAACTCCTGTGTCCMAIGG 6125
 QY 5160 tctccaccactcaccatcacctgctgtagcaagagtccttagtcaggggaggtgcatttt 5219
 Db 6126 TCTCCACCACCTACCATCACCTGCTGTAGCAAGAGTCCCTAGTCAGGGAGGTGCAATTTT 6185
 QY 5220 agtagttaattgcacttatccatgagataaataaaaggaggaactgtttttatcagtcgga 5279
 Db 6186 AGTAGTACATTGCACCTTATCCATGAGATAAATAAAGAGAVCTGTTTTTATCAGTGGA 6245
 QY 5280 ggttaacctaaaatttcaaaagtgcctttttgaaatcttgggctcctctctctctaga 5339
 Db 6246 GGTACCTAAATTTCAAAAGTGTGCGCTTTTGAATCTTTGGGCTCTCTCTCTGTAGA 6305
 QY 5340 accaatggcccttgggtcacggcctgcacactcaactgaagtagtctctagctcctgca 5399
 Db 6306 ACCAATGCCCTTTGTGGCTCAGGGCTCGCACCTAACTGGAGAGTTCTGAGCTCCTGCA 6365
 QY 5400 gctcacctgagccacagactaggtcttggctccttccgc 5441
 Db 6366 GCTCAGCTGAGCCACAGACTAGGCTTCTTGGCTCCTTCCGC 6407

RESULT 2

US-08-599-654-7
 ; Sequence 7, Application US/08599654
 ; Patent No. 5882925
 ; GENERAL INFORMATION:
 ; APPLICANT: FALB, DEAN A
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 ; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
 ; NUMBER OF SEQUENCES: 54
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PENNIE & EDMONDS
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/599,654
 ; FILING DATE: 09-FEB-1996
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/485,573
 ; FILING DATE: 07-JUN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/386,844
 ; FILING DATE: 10-FEB-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: CORUZZI, LAURA A
 ; REGISTRATION NUMBER: 30,742
 ; REFERENCE/DOCKET NUMBER: 7853-041
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-8864
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 6407 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: both
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: CDNA
 ; US-08-599-654-7

Query Match	73.3%;	Score	5098.4;	DB 2;	Length	6407;			
Best Local Similarity	94.3%;	Pred. No.	0;						
Matches	5412;	Conservative	21;	Mismatches	8;	Indels	301;	Gaps	2;

QY	1	gttcgatgaagaattgcgcgtttttcaacaaagaagtgaacagagctcggagatgggaac	60
Db	666	gttcgatgaagaattgcgcgtttttcaacaaagaagtgaacagagctcggagatgggaac	725
QY	61	agagagggcgatggggctgtcagaagaatgactgtgcacagccaagggccaccacttc	120
Db	726	agagagggcgatggggctgtcagaagaatgactgtgcacagccaagggccaccacttc	785
QY	121	ggcttggagcccttccttttctcgtgtttggagatgggagctgacacagccttctag	180
Db	786	ggcttggagcccttccttttctcgtgtttggagatgggagctgacacagccttctag	845
QY	181	gaagagaattctcagagaccagatctcctcgtcgtcatttctacagagacagacttc	240
Db	846	gaagagaatttccctcagagaccagatctcctcgtcgtcatttctacagagacagacttc	905
QY	241	ctctcctcttagacctttcctcactcttgaagatcacagaagcttaacaactccac	300
Db	906	ctctcctcttagacctttcctcactcttgaagatcacagaagcttaacaactccac	965
QY	301	tggcctccagagctcctcagtcagtcacaaagaacatgcattgtgtacacgtgtccac	360
Db	966	tggcctccagagctcctcagtcagtcacaaagaacatgcattgtgtacacgtgtccac	1025
QY	361	tgatgtggcccgagaacgctgcgacttttgacggtcagctcgtggacctgtgagcaaac	420
Db	1026	tgatgtggcccgagaacgctgcgacttttgacggtcagctcgtggacctgtgagcaaac	1085
QY	421	agaaggtctcccaaggactccagaattgccaagactccactcagtcctctcttcacc	480
Db	1086	agaaggtctcccaaggactccagaattgccaagactccactcagtcctctcttcacc	1145
QY	481	ctctgcagtggaatcgagaagaacagtagagtaactgggaatccagggatggagaatt	540
Db	1146	ctctgcagtggaatcgagaagaacagtagagtaactgggaatccagggatggagaatt	1205
QY	541	catggaacatccacagaaaaatgaattggacttaactctttgag - tggcaaaatgattc	599
Db	1206	catggaacatccacagaaaaatgaattggacttaactctttgag - tggcaaaatgattc	1265
QY	600	cccaaccttgagaacaacatcagcttgcagcagctctgaggtgcaaaatggaagtcctac	659
Db	1266	cccaaccttgagaacaacatcagcttgcagcagctctgaggtgcaaaatggaagtcctac	1325
QY	660	gtctcagactgagactgtgtctagtcagtcgacccatgagagtgagagatcaactgc	719
Db	1326	gtctcagactgagactgtgtctagtcagtcgacccatgagagtgagagatcaactgc	1385
QY	720	acaatggctcttgaccaacagcaacaatctgcagatgtgcaggaagctcgtgctcata	779
Db	1386	acaatggctcttgaccaacagcaacaatctgcagatgtgcaggaagctcgtgctcata	1445
QY	780	tccctgaagtgatgaactcagctgagtgagaccagttctcagactcactgctacagctcgg	839
Db	1446	tccctgaagtgatgaactcagctgagtgagaccagttctcagactcactgctacagctcgg	1505
QY	840	aggaagtccacagcattgggagatagaggttatccagagttcttcactatcattctctc	899
Db	1506	aggaagtccacagcattgggagatagaggttatccagagttcttcactatcattctctc	1565
QY	900	ggaaagcttgaattcatcagcaccacgctgggagaacgttcaa - - - - -	940
Db	1566	ggaaagcttgaattcatcagcaccacgctgggagaacgttcaa - - - - -	1625
QY	941	- - - - -	940
Db	1626	gccaggccaaagcactagtgtagcttcggccaatcgacagagcagaggacttctgggggtgcc	1685

QY 1800 tcaatccttaacagagtcctccacagcagcaaacctctccagccacagcaccacttagc 1859
Dd 2766 TCCATCCTTAACAGAGTCCTCCACAGAGCAAAACCTTCCAGCCACAAGCAACCTTAGC 2825
QY 1860 acaaatgtctccaaattcaacaactaccattctgaagaacctctcagcctcttatgaccac 1919
Dd 2826 ACAAAATGTCFCCAACTTTCAAACTACCAATTCFCAAGAGCCTCTCAGCCCTTTATGACCAC 2885
QY 1920 tcttgccacctgtcaagcacagcatctctgttcaactggcccttatagccgtacagactac 1979
Dd 2886 TCCTGGCACCTGTCAAGCACACATCTCTGGTCACCTGGCCCTATAGCCGTACAGACTAC 2945
QY 1980 agtggaaaacagctctcgtgacccatcctgtaaa tactagttctcctaattcacaaga 2039
Dd 2946 AGTGGAAAACAGCTCTCGTGTACCCATCCTGAAATACTAGTTCCCTCAAAATCTCAACAGA 3005
QY 2040 agtggcatcagcacagaagaacagagattgttgatgtctaccacgtgattgattccc 2099
Dd 3006 AGTGGCATCAGCACAGAAGGAACCGAGTGATTGTGGATGCTTACCACCTGGATTGATCCC 3065
QY 2100 ttgaccaggtgtaccacacacagcaaaagaaaagcacaaaagcttggggttacagcaga 2159
Dd 3066 TTTGACCAAGTGTACCCACATCAGCAAAAGAAATGACCAAAAGCTTGGCGTTACAGCAGA 3125
QY 2160 gtacagccagcttcacgttccctcgtggaacatctcctctcccaaacacagttgtttc 2219
Dd 3126 GTACGCCAGCTTCAGTTTCCCTCGGAACATCTCCTTCTCCCAAAACCAACAGATTGTTTC 3185
QY 2220 cagggctgaagacttggtctcccaaatctccaccttctgttccagcagacacagtc 2279
Dd 3186 CAGGCTGAAGACTTGGCTCCCAATCTGCCAATCTGCCAATCTTGTGTTTCAGAGCAGCACAGTC 3345
QY 2280 accaaacaaactgtcctcttcagccctcagtcacacagctgtgctgtgaaccttgttctca 2339
Dd 3246 ACCAAACAACACTGCTCTTTCAGCCTCAGTCAACAGCTGTGCTGTGAACCTTGTCTTCA 3305
QY 2340 caatggcgaaatggctgcgagacaaacccagccgtgctaccactgaggtgcgcgccttc 2399
Dd 3306 CAATGGCGAATGGCTGCAGACAAACACAGCCGTGCTTACCAGTGAAGTGGCCGCTTC 3365
QY 2400 ctggcagggagtgattgcaagtggatgtgaatgagtgagctgcgaacccctgcacatc 2459
Dd 3366 CTGGCAAGGGATGATGCAAGTGGATGTGAATGAGTGCCTGTCCGACCCCTGCCCATC 3425
QY 2460 cacagccactgtcaacaa tactcagggatccttatactgcaaatgcccgggtgggtacca 2519
Dd 3426 CACAGCCACGTGCAACAATACTCAGGGATCCTTTATCTGCAAAATGCCGCTTGGGTACCA 3485
QY 2520 gttggaaaaagatatgcaatttggtagaaccttcgtgacagagtttaaatataagag 2579
Dd 3486 GTTGGAAAAAGGGATATGCAAAATTTGGTTAGAACCTTCGTGACAGAGTTTAAATTAAGAG 3545
QY 2580 aactttcttaaTacaactgtgaaaaacattcagacctacagaagagttgaaaatgagat 2639
Dd 3546 AACTTTTCTTAATACAACTGTGGAANAACATTCAGACCTTACAGAAGTTGAAATGAGAT 3605
QY 2640 caccaaaacgttaaatatgttttttcagcgttacctagttacatccgatctcacagtcca 2699
Dd 3606 CACCAAAACGTTAAATATGTGTTTTTCAGCGTTACCTAGTTACATCCGATCTACAGTTCA 3665
QY 2700 cgctctagggagtcacaagcgggtgtgatatctcaactgcaaaacaccttttccctggcctc 2759
Dd 3666 CGCCTCTAGGGAGTCCAACCGGTGGTGTATCTCACTGCAAAACAACCTTTTCCCTGGCCTC 3725
QY 2760 caatggagcctatttgacctgctataggatgagaaaatgtgtcaactcctgcgaatc 2819
Dd 3726 CAATGTGACGCTATTTGACCTGGCTGATAGGATGCAAGAAATGTCAACFTCCCTGCAAGTC 3785
QY 2820 ctctgctgaggtctgcagctctctgggattcagagcggagatcttttagagcgggcagctt 2879
Dd 3786 CTCTGCTGAGGTCTGCCAGTCTTGGGATCTCAGAGCGGATCTTTAGAGCGGCGACCTT 3845
QY 2880 gtgcaagcggaagagtcctcogaaatgtgacaaaagacacctccatctgcactgacctggacgg 2939

Dd 3846 GTCAACGCGAAAGAGTCCCGAATGTGACAAAGACACTTCCATCTGCACCTGACCTGGACGG 3905
QY 2940 cgttccctcgtgccagtcaagtccggaacttctcagttcaacaagatgaccactctg 2999
Dd 3906 CGTGGCCCTGTGCCAGTGCNAAGTCGGGATACTTTTCTAGTTCAACAAGATGACCACTCCTG 3965
QY 3000 ccgagcatgtgaagatggatataggtctgaaatgaaacctgcattgattggccattgg 3059
Dd 3966 CCGAGCATGTGAAGATGGATATAGGCTTGAATGAAACCTGCATGAGTTGCCCATTTGG 4025
QY 3060 ccttggtgtctcaactgtggaaaacccctatacagcttatacactgtggtgagtcagccgc 3119
Dd 4026 CCTTGTGTGTCTCAACTGTGAAACCCCTATCAGCTTATCACTGTGTGTATCCGACCCGC 4085
QY 3120 gggagtggtgctcctcctcactcctaggtcgcactgattgttacctgttgccagaagaa 3179
Dd 4086 GGGAGTGGGCTCCTGCTCATCTAGGCATCGCACTGATGTTTACCTGTTGAGANAAGAA 4145
QY 3180 taataatgacataagcaaacctcattctcaaaagtggagatttcccaatgtccccaatgc 3239
Dd 4146 TAAAAATGACATTAAGCAAACTCATCTTCAAAAGTGGAGATTTCCAAATGTCCCATATGC 4205
QY 3240 tgaatcccccaaaaaatcctogtccaaagaatggggccgagagaagctattgaaatgcata 3299
Dd 4206 TGAATACCCCAAAAAATCCTCGCTCACAAAGATGGGCGGAGAGCTATTGAAATGCAATGA 4265
QY 3300 gaatgaagtaccaaaaacccctcctccagatgcaggtatgtgtactactcgcctcacaagt 3359
Dd 4266 GAATGAAGTACCANAACCTCTCCAGATGACGGATGTGTACTACTCGCTACAAAGTGT 4325
QY 3360 aaggaatccagaacttgaaacgaacgcactctaccggcctacactggagctgcaggaatc 3419
Dd 4326 AAGGAATCCAGAACTTGAACGAACGACTCTACCCGGCTTACACTGGACTGCGCAGGATC 4385
QY 3420 acggcaattctgattttcccgagacagtataaacctgttctcactcagtgatgaaagcag 3479
Dd 4386 ACGGCATCTGTGATTTTCCCGGACAGTATAACCCGCTCTTTCATCAGTGTGAAGACAG 4445
QY 3480 aagaagagactacttttaagtcacagagagagagagactcattgctcgtgagccagtcacc 3539
Dd 4446 AACAAGAGACTACTTTTAAGTCCAGGAGAGAGAGGACTCATTTGCTGTGAGCAGTCAAC 4505
QY 3540 tgggacctctgtcagagagaccacacagagagctgcgcccagattgttcggagagccac 3599
Dd 4506 TGGGACCTCTGCTCAGAGAGCCGACAGAGGCTGCGCCGAGATTTGCGGGAGGCAC 4565
QY 3600 gctgagtggaagcagagagagagagcagcatcggggctgaccacagtgaggagagaca 3659
Dd 4566 GCTGAGTGGCAACAGGAAGAGGACAGGCTATCGGGGCGGTGACCAGTGGAGGAGACA 4625
QY 3660 ggtggatgtggaaccacaggtcctcattcagcacctttgttactgctgaaactgaaat 3719
Dd 4626 GGTGGATGTGGAACCAACAGCTGCTCATTCATCAGCACCTTTGTTACTGTGGAAGTGAAT 4685
QY 3720 gtgggcagtatcaagagagctctctctgagtcgacctgagccatggcactggcagagggcg 3779
Dd 4686 GTGGCCAGTATCAAGAGAGTCTCTCTGAGTACTGCACCATGCACTGCGACCTGCGCAGGCGC 4745
QY 3780 actattagccagggcagacacactagacttcagtcgagggacctggtttcccttcgtttg 3839
Dd 4746 ACTATTAGCCAGGGCAGACCACTAGACTTTCAGTGCAGGAGCTGGTTTCCCTTCGTTTG 4805
QY 3840 cacttagtaaaattgggtgggaggtttcccttttggatctgttttgagactgtttccagaaa 3899
Dd 4806 CACTTTAGTAAATTTGGTGGGAGGTTTCCCTTTGGATCTGTTTGGACATGTTTCCAGAAA 4865
QY 3900 gaaggtctccttcccgagacacttccatagggcagcaatttggatgatttcatttgcagcaa 3959
Dd 4866 GAAGGCTTCTTCCCGAGACACTTCCATAGGCAGCAAAATTTGGTGAATTCATTTTGCASCAA 4925
QY 3960 aatactggctgttaattatttctcctgccagcgctgctgcttaaacacacagatgagga 4019

Db	4926	AAATCTGGCTGTGTTAAATATATTTTCTTGCCCGACGRCCTCGGTGCTAAACACACAGATGAGGA	4981
Qy	4020	tgagcgtaccactgaagctgaagatgctgcacctgaacgacacagtgcttttcatactgttt	4079
Db	4986	TGASCGTACCACCTGAAGTCTGAAGATGTGCGCAATTGAACGGACAGTGTTTCATATGTTT	5045
Qy	4080	ctagggtgttattgctacagttctcaagccagccccacacagtgtagaataatgtgtgaggc	4139
Db	5046	CTAGGTGTCTTATGTCTACAGTTTCCAAAGCCASCCCCACACAGTGAAGAAATGCTGAGGC	5105
Qy	4140	accgcacacaactgcaatggttttttaagtcaagtgacacatgattttaagatttttt	4199
Db	5106	ACCGCACAACTGCAATGTGTTTTTAAGTCAAGGTGACACATGATATTTAAGATTTTTT	5165
Qy	4200	tttaaaactcttgcagttaaactcacttttcaaaagcctggatcagggcacaac	4259
Db	5166	TTTAAATCTYTTGCAGTTAAATCTCACTTYYTCAACAGCCTGGATCAGGCCAATAC	5225
Qy	4260	aaettatattgggttttagctggagctcagcaggcagattgcaggcagggggcaccttt	4319
Db	5226	AACTTATATVTGTTTTAGTCTGGAGCTCAGCAGCAGATGTCAGGCAGGGGGCACTTT	5285
Qy	4320	tcatcatgagggccagcctgggcctgggactctgataccactgtgagggccaggg	4379
Db	5286	TCATTCATGAGGGCCACGCTGGGCGCTGGGACTCTGATACCAATGTGGAGGCCAGAGG	5345
Qy	4380	cagctcgctatgagagagaatgtcaaaactgaacacaggtttcaccactctagaagaaca	4439
Db	5346	CAMCTGGTATGAGGAGGAANTGTCAACTGAACCGCAGGTTTCACCACTCTAGGAAGACA	5405
Qy	4440	gcttgttgagccctcagctggatgtggttagagggatggcctgaaataggcaggtttaga	4499
Db	5406	GCTTGTGTGACCCCTCCASCTGSGATGTGTTAGAGGGATGGCTGAATAGSCAGTTAGA	5465
Qy	4500	ttctcgtcatcaacagctgtttggagaactgtgtggaatcctcagggaagaacaggagcc	4559
Db	5466	TTTTCTGTCATCAACAGTGTGTTTGGGAASTCTGTGGATTCTCTGAGGAAGAACAGGGAGCC	5525
Qy	4560	gagatggagccacacatgattgtctaccggctactcacacacttgtaccacagaatct	4619
Db	5526	GAGATGGAGCCACATGAATTTGCTACCGGCTACTCGACGACTTGTACCCAGAAATCT	5585
Qy	4620	catgtccacaaaccccatgtaaactttcaacactcaagctgtttattcgcgctgaagaa	4679
Db	5586	CATGTCACAAACCCCATGTAAACTTCAACCACATCAAGSTGTTTATTCGGCTGAAGAA	5645
Qy	4680	ataactttttttctcaccagctcattgtgaactcttcataatgctgtgtgcacacctcc	4739
Db	5646	ATAACTTTTTTTTTCTCACCCAGTCAATTTGTACTCTTCTCATATGGSTATGTGCGACCCCTCC	5705
Qy	4740	agaaacgtgttatacttcagctcagttggaagaaactgaagactccggttgtgtcaggg	4799
Db	5706	AGAAACGTGGTTATCTKCAATGCTAGTGTGGGAGAACTTGAAGACTTCGGTGTGTCGAGG	5765
Qy	4800	aactgaggggttgacctcgggaaggaagtccacactcttattattatgtcctgtgatg	4859
Db	5766	AACTGAGGGTTGACCTTCGGGAAGGAAGTTCCACATCATCTTATTTATTTATGSCCTGTGATG	5825
Qy	4860	tgggtcctgccaggagagacatccagttactcgggtgtctttaatgccacctggggaactgt	4919
Db	5826	TGGGTCTCTGCCAGGGAGACATCCAGTACTCGGTGTCTTAAATTTGCCACCTTGGGGAAGCTGT	5885
Qy	4920	gtttattggcctctcttggggcatcctggttttgatgaagtgaaggaatacacagagta	4979
Db	5886	GTTTTATTGGCCTTCTTTGGGGCATCTCGTTCGGATGAAGTGAAGGAATACAGAGGTA	5945
Qy	4980	aaagaattgtctcacacctgaagcggggagttccgccttcaacttcttggaatgggtcag	5039
Db	5946	AAAGAATTGTCTCCACCCCTGAACGGGGAGTCCCGCTTCACATTTCTGGAAATGGTCAG	6005
Qy	5040	ccactggggacagttctgccccgggcatgtgttttctcagggtctctctaaataataatc	5099
Db	6006	CCACTGGGACAGTCTGCCCCGGCATGGTTGTCTTCAAGTCTCTAAATATATATC	6065

RESULT 3
 US-08-944-868A-7
 ; Sequence 7, Application US/08944868A
 ; Patent No. 6018025
 ; GENERAL INFORMATION:
 ; APPLICANT: FALB, DEAN A
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 ; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
 ; NUMBER OF SEQUENCES: 54
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PENNIE & EDMONDS
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/944,868A
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/599,654
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/386,844
 ; FILING DATE: 10-FEB-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: CORUZZI, LAURA A
 ; REGISTRATION NUMBER: 30,742
 ; REFERENCE/DOCKET NUMBER: 7853-041
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-8864
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 6407 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: both
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: cdna
 US-08-944-868A-7

Query Match 73.3%; Score 5098.4; DB 3; Length 6407;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 5412; Conservative 21; Mismatches 8; Indels 301; Gaps 2;

Qy 1 gttcgatgaagaattgcgctttttcaaaacaaagtggaacagcctcgggatgggaac 60
Db 666 GTTCGATGAAGAATGTCGGCTTTTCAAAACAAGAGTGAACAGCCTCGGAGATGGGAAC 725

Qy 61 agagagggcgatggggctgctcagaagaatgactgtgcacagccaaagagccaccacttc 120
Db 726 AGAGAGGGCGATGGGGCTGCTCAGAGAATGACTGTGCACAGCCAAAGAGGCCACACATTC 785

Qy 121 ggcctggagccctctcttctctctgttggagatggagagctgacacagcctcttag 180
Db 786 GGCCTGGAGCCGCTCTTCTCTCTGCTTTGGAGATGGAGAGTGCACACGCTTCTAG 845

Qy 181 gaagagaaattcctcaggacagatctctctggtgctggaattctacaggacagcttc 240
Db 846 GAAGAGAAATTCCTCAGGACCAGATCTCTCTGCTGCATTTCTACAGGACAGCAGCTTC 905

Qy 241 ctctctctcttagaccttctcaccctctgaaagtcagagagcttaacaaactccac 300
Db 906 CTCCTCTCTTAGACCTTCTCACCCTCTGAAAGTACAGAGAGCTTAACAATCCAC 965

Qy 301 tggcctccagagctcctcagtcagtcacaaacaaagacaaatgcatgttgcctacgtgttcac 360
Db 966 TGGCCTCCAGAGCTCCTCAGTCAGTCAAAACAAGACAATGCATGTTGCTACCGTTCAC 1025

Qy 361 tgatggtggccggaacgctgcatctcttgaagtcagagagcttggaacctgtgaacagac 420
Db 1026 TGATGGTGGCCGAGAGCGCTGCGCATCTTTGACGGTCACTGCGGACCTGTGAGCAAGAC 1085

Qy 421 agaaggtctcccaaggactccagaattggccagacttcaacctcagctctctcttacc 480
Db 1086 AGAAGGCTTCCCAAGGACTCCAGAAATGGCCAGCCTTCATCTCAGTCTCTTTTACC 1145

Qy 481 ctctgcagtggaatcgaagaacacagtagagtaactgggaatccaggggatgaggaatt 540
Db 1146 CTCGCACTGGAATCGAGAGAAACAGTAGAGTAACCTGGGAATCCAGGGGATGAGGAAT 1205

Qy 541 catgaaccatccacagaaatgaatttgagcttaagctcttggg -tggaacaaatgattc 599
Db 1206 CATTAACCATCCACAGAAATGAATTTGGACTTACGCTCTTTGGCTTGGCAAAATGATTC 1265

Qy 600 cccaacctttgagaaacatcagcttgcagcagctctgaggtgcacaaatggaagtcacat 659
Db 1266 CCCAACCTTTGGAGAACATCAGCTTGGCAGCAGCTCTGAGGTGCAAAATGGAAATCCCAT 1325

Qy 660 gtctcagactgagactgtgtctaggtcagtcgcacccatgagaggtggagagatcactgc 719
Db 1326 GTCTCAGACTGAGACTGTGTAGGTGAGTCGCACGCCATGAGAGGTGGAGAGATCACTGC 1385

Qy 720 aactggtctctgccaacagcaacaacatctgcagatgtgacaggaagctctgttcata 779
Db 1386 ACATGTGCTCTTGACCAACAGCACAAATCTGCAGATGTGCAGGAAGCTCTGTCTTCATA 1445

Qy 780 tcttgaaggtgtaagtcttcagtggttgacccagcttctcagactcctactatacagctgg 839
Db 1446 TCCTGAAGGTGTAAGTCTTCAGTGTGACCCAGTCTCAGACTCTACTGTACAGTCTGG 1505

Qy 840 aggaagtacacagcattgggagataggagtatttcagagcttctcatctacatcttctc 899
Db 1506 AGGAAGTCACACAGCATTTGGAGATAGGAGTTATTCAGAGTCTTTCATCTACATCTTCTC 1565

Qy 900 ggaagcttgaaattcatcagcaccacgtggagaacgttcaa - 940
Db 1566 GGAAGGCTTGAATTCATCAGCACACACGTGGAGAACGTTCAACCTTTGGAAGACAGCCGAGA 1625

Qy 941 ----- 940
Db 1626 GCCAGGCCAAGCACTAGGTACAGTTCCGCCAATGCAGAGACAGGACTTCCTGGGGTGCC 1685

Qy 941 ----- 940
Db 1686 CTCCTCGGACCCACACACCTTGGCTACTGTCTACTGGAACGGGGAACGCACACTCGGGTC 1745

Qy 941 ----- 940
Db 1746 TGTCAACCCTCACCAACACCAGCATGAGCAGCACTTCTGGGAAGCAGCGCCCTGCAGC 1805

Qy 941 ----- 940
Db 1806 GGCCATGCCCCCAAGAAACAGAGGGTGCCCTCTCTGCACCTAAACGTAGCGGACGACATGGG 1865

Qy 941 -----tcgctgggattagctacgg 959
Db 1866 CTTGGTCTCACGGTCACTGGCCGCTCCAGTGCACCTCGAGTGCCTGGGATTAGCTACGG 1925

Qy 960 tcaagtgctggcagagctattgaacaaagagacttccagcagccacacacagaccaccta 1019
Db 1926 TCAAGTGGTGGCAGAGCTATTGAACAAAGGACTTCCAGCGACCAACACAGACACACCTA 1985

Qy 1020 cctgtcactacttccaccaagagagcggcgttactgtccattacagataacagttc 1079
Db 1986 CTTGTCTACTTTTCCAAAGGAGAACGGCGGTTACTGTCCATTACAGATAACAGTTC 2045

Qy 1080 atcctcagacattgtggagagctcaactcttattataaaatctcaaaactcttcacattc 1139
Db 2046 ATCCTCAGACATTGTGGAGAGCTCAACTTCTTATATTAATAATCTCAAACTCTTCACATTC 2105

Qy 1140 agagtattcctcctcttctctatgtcagactgagagaagtaaatcctcactatgacgg 1199
Db 2106 AGAGTATTTCCTCCCTTTTCTCATGTCTCAGACTGAGAGAGTAACATCTCATCTATGACGG 2165

Qy 1200 ggaatagctcagccttctactagtcgagccagttctgcatacaccaccccttcgctcta 1259
Db 2166 GGAATATGCTCAGCCTTCTACTGAGTCGCCAGTTCTGCTATACATCCAACTCTCCGTCCTA 2225

Qy 1260 cacaccacacataaatatgcgaacaccttcggtgtgtctgacactgatgctgagttgt 1319
Db 2226 CACACCACCACTTAATATGCGGAACACTTCGGTGTGTCTGCACACTGATGCTGAGTTGT 2285

Qy 1320 tagtgaactcctcctcctctctctcctcctcctcctcctcctcctcctcctcctcctt 1379
Db 2286 TAGTGACTCTCTCCCTCCCT 2345

Qy 1380 gctctgcccctctgtgcacaaatcccacatttatttcaatatttaccataaacag 1439
Db 2346 GCCTCTGCCCTCTCTGTGCACAAATCCCACCAATTTATTTTCATCAATTTTACCATCAACGAG 2405

Qy 1440 ggcctctgtgcatactactaaagtctacctctgtatgcataccacacatgggtcttctcacc 1499
Db 2406 GGCCTCTGTGCATCTACTAAAGTCTACCTCTGATGCATCCACACCACTGGTCTTCTCTCACC 2465

Qy 1500 ataccctttaccagatactcttaacagacatctacatctcgcacacatttctgtctcacaac 1559
Db 2466 ATCACTTTTACCAGATATCTTAACGACATCTACATCTGCCCCACTTTCTGTCTCACAAAC 2525

Qy 1560 aacctgtccacagtcacttctacccctctcctcgcagggcagggagagactcctgtgac 1619
Db 2526 AACCTTGCCACAGTCATCTTCTACCCCTGTCTCCCGCAGGGCAAGGGAGACTCTCTGTGAC 2585

Qy 1620 ttcaatttcagacatacaacaaatgacatctatgcataatgcataatgtagtagtaacactgc 1679
Db 2586 TTCATTTTCAGACATCAACAATGACATCATTCATGCACAAATGCTCCATAGTAGTCAAACTGC 2645

Qy 1680 agaccttaagagccagagacccccacacccaagaagaagtattacagaaatcaaaagtacac 1739
Db 2646 AGACCTTTAAGGCCAGAGACACCCACACCAAGAGAGAAAGTCAATTACAGAAATCAAAAGTCAAC 2705

Qy 1740 aagcctgtgtctctcgcacagagtcacccaagctgttaacaaactctccttggc 1799
Db 2706 AAGCCTGTGTCTCTGCCCCACAGAGTCCACCAAGAGCTGTAAACAACAACACTCTCTCTTTGCC 2765

Db 4926 AATACTGGCTGTGTAATATTTTCTCCAGCGRCTCGCTGCTAAACAACATGAGGA 4985
QY 4020 TGAGCTACCACTGAAGTCTGAAGATGCTGCCATTGAACGACAGTGTCTTCCATATGTTT 4079
Db 4986 TGASCGTACCACGAACTCTGAAGATGCTGCCATTGAACGACAGTGTCTTCCATATGTTT 5045
QY 4080 CTAGGTTGTTCTTATGCTACAGTCTTCCAAAGCCASCCCCACAGTGAGGAGGC 5105
Db 5046 CTAGGTTGTTCTTATGCTACAGTCTTCCAAAGCCASCCCCACAGTGAGGAGGC 5105
QY 4140 ACAGCACAACTGCAATGTGTTTTTAAAGTGAAGGTGACACATGTATTTAAAGATTTTTT 5165
Db 5106 ACCGCACAACTGCAATGTGTTTTTAAAGTGAAGGTGACACATGTATTTAAAGATTTTTT 5165
QY 4200 TTAAATATCTCTGAGTGAAGTCTTCAAGGTGACACATGTATTTAAAGATTTTTT 5165
Db 5166 TTAAATATCTCTGAGTGAAGTCTTCAAGGTGACACATGTATTTAAAGATTTTTT 5165
QY 4260 AACTTATATGTTTGTAGCTGAGGCTCAGCAGGAGATTCAGGAGGAGGAGGAGGAGG 5285
Db 5226 AACTTATATGTTTGTAGCTGAGGCTCAGCAGGAGATTCAGGAGGAGGAGGAGGAGG 5285
QY 4320 TCAATCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5345
Db 5286 TCAATCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5345
QY 4380 CAGCTGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5439
Db 5346 CAMCTGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5439
QY 4440 GCTTGTGACCCCTGCACTGAGTGTGTTAGAGGAGTGGTGTAGAGGAGTGGTGTAGAG 5465
Db 5406 GCTTGTGACCCCTGCACTGAGTGTGTTAGAGGAGTGGTGTAGAGGAGTGGTGTAGAG 5465
QY 4500 TTCTGCTCACTCAACAGTGTGTTGGAAGTGTGTTGGAAGTGTGTTGGAAGTGTGTT 5499
Db 5466 TTCTGCTCACTCAACAGTGTGTTGGAAGTGTGTTGGAAGTGTGTTGGAAGTGTGTT 5499
QY 4560 GAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5525
Db 5526 GAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5525
QY 4620 CATGTCACAAACCCCATGTAATTTTCAACACTCAACAGTGTGTTTATTCGGCTGAAGAA 5645
Db 5586 CATGTCACAAACCCCATGTAATTTTCAACACTCAACAGTGTGTTTATTCGGCTGAAGAA 5645
QY 4680 ATAACCTTTTCTTCAACAGTGTGTTGGAAGTGTGTTGGAAGTGTGTTGGAAGTGTGTT 5705
Db 5646 ATAACCTTTTCTTCAACAGTGTGTTGGAAGTGTGTTGGAAGTGTGTTGGAAGTGTGTT 5705
QY 4740 AGAAGCTGTTTCTTCAACAGTGTGTTGGAAGTGTGTTGGAAGTGTGTTGGAAGTGTGTT 5799
Db 5706 AGAAGCTGTTTCTTCAACAGTGTGTTGGAAGTGTGTTGGAAGTGTGTTGGAAGTGTGTT 5799
QY 4800 AACTGAGGTTGACCTTCGGAAGGAGTTCACCTCATCTTATTTATTTATGCTGTGATG 5825
Db 5766 AACTGAGGTTGACCTTCGGAAGGAGTTCACCTCATCTTATTTATTTATGCTGTGATG 5825
QY 4860 TGAGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5885
Db 5826 TGAGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5885
QY 4920 GTTATGAGGCTTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5979
Db 5886 GTTATGAGGCTTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5979
QY 4980 AAGAAATGCTCCACCTCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6005
Db 5946 AAGAAATGCTCCACCTCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6005
QY 5040 CCACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6099
Db 6006 CCACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6099

QY 5100 CCTATTCTTACATAAATCTTGGCCCTGATGGTTTAAAGCAAGAACTCCTGTGTCCTCATGG 5159
Db 6066 CCTATTCTTACATAAATCTTGGCCCTGATGGTTTAAAGCAAGAACTCCTGTGTCCTCATGG 6125
QY 5160 TCTCCACCACTCACCATCACCTTCTGTAGCAAGAGTCTTAGTCAGGAGGAGTGTGTTTATCAGTGA 5219
Db 6126 TCTCCACCACTCACCATCACCTTCTGTAGCAAGAGTCTTAGTCAGGAGGAGTGTGTTTATCAGTGA 6185
QY 5220 AGTAGTTAAATTCACCTTATCCATGAGATAAATAAAGAGAGTGTGTTTATCAGTGA 5279
Db 6186 AGTAGTTAAATTCACCTTATCCATGAGATAAATAAAGAGAGTGTGTTTATCAGTGA 6245
QY 5280 GGTAACTTAAATTCACCTTATCCATGAGATAAATAAAGAGAGTGTGTTTATCAGTGA 5339
Db 6246 GGTAACTTAAATTCACCTTATCCATGAGATAAATAAAGAGAGTGTGTTTATCAGTGA 6305
QY 5340 GGTAACTTAAATTCACCTTATCCATGAGATAAATAAAGAGAGTGTGTTTATCAGTGA 5399
Db 6306 GGTAACTTAAATTCACCTTATCCATGAGATAAATAAAGAGAGTGTGTTTATCAGTGA 6365
QY 5400 GGTAACTTAAATTCACCTTATCCATGAGATAAATAAAGAGAGTGTGTTTATCAGTGA 5441
Db 6366 GGTAACTTAAATTCACCTTATCCATGAGATAAATAAAGAGAGTGTGTTTATCAGTGA 6407

RESULT 4
US-08-944-423A-7
; Sequence 7, Application US/08944423A
; Patent No. 6020463
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITL OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: WINDOWS 95
; SOFTWARE: PatenCin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,423A
; FILING DATE: 06-OCT-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: JUN-07-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEFAX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6407 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both

; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
US-08-944-423A-7

Query Match 73.3%; Score 5098.4; DB 3; Length 6407;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 5412; Conservative 21; Mismatches 8; Indels 301;

Qy	1	gttcgatgaagaattgcgcgcttttcaacaagaagtggaacagcctcgtgagatgggaac	60
Db	666	gtttcggatgaagaagaatttgcgcgcttttcaaaacaagaagtggaacagcctcgtgagatgggaac	725
Qy	61	agagaggcgatgggctgtcagaagaatggactgtcacagccaagagccacacattc	120
Db	726	agagaggcgatgggctgtcagaagaatggactgtgcacccaagagccacacattc	785
Qy	121	ggcttggagcccttccttctctgtcttggagatggagagctgaccacgcctctctag	180
Db	786	ggcttggagcccgctcttcttcttcttggagatggagagctgaccacgcctctctag	845
Qy	181	gaagagaaattctcagggaccagatctctctggctgcatcttctcagagacagcattc	240
Db	846	gaagagaaattctcagggaccagatctctctggctgcatcttctcagagacagcattc	905
Qy	241	ctctcctctctagaaccttctcaccctctgaagatcacagagaagcttaacaactcac	300
Db	906	ctctcctctctagaaccttctcaccctctgaagatcacagagaagcttaacaactcac	965
Qy	301	tggcctccagagctctctcagtcagtcacaaacaagaacaatgcatgtgtctaccggttccac	360
Db	966	tggcctccagagctctctcagtcagtcacaaacaagaacaatgcatgtgtctaccggttccac	1025
Qy	361	tgatgtggcccgagaaacgctcgatctttgacgctcagtcgtggacactgtgagcaagac	420
Db	1026	tgatgtggcccgagaaacgctcgatctttgacgctcagtcgtggacactgtgagcaagac	1085
Qy	421	agaagggcttcccaagagactccagaaattgccacgacttcactcagtcctctcttccacc	480
Db	1086	agaagggcttcccaagagactccagaaattgccacgacttcactcagtcctctcttccacc	1145
Qy	481	ctctgcagtggaatcagagaagaacagtcagatgaactgggaatccaggggatgaggaatt	540
Db	1146	ctctgcagtggaatcagagaagaacagtcagatgaactgggaatccaggggatgaggaatt	1205
Qy	541	cattgaacatccacagaataatgaattggacctcagctcttggcgtggcgaataatgaggaatt	599
Db	1206	cattgaacatccacagaataatgaattggacctcagctcttggcgtggcgaataatgaggaatt	1265
Qy	600	cccaaccttggagaacatcagcttgcacgacgctcctgaggtgcataatggaagcccat	659
Db	1266	cccaaccttggagaacatcagcttgcacgacgctcctgaggtgcataatggaagcccat	1325
Qy	660	gtctcagactgagactgtcttagtgcagtcgccccatgagagtgagagatcactgc	719
Db	1326	gtctcagactgagactgtcttagtgcagtcgccccatgagagtgagagatcactgc	1385
Qy	720	acactggctcttgaccaacagacacaactctgcagatgtgacaggaaactctgcttcaata	779
Db	1386	acactggctcttgaccaacagacacaactctgcagatgtgacaggaaactctgcttcaata	1445
Qy	780	tcttgaaggtgaaatgcttctcaggttgaccagcttctcagactctactgtcacagctcg	839
Db	1446	tcttgaaggtgaaatgcttctcaggttgaccagcttctcagactctactgtcacagctcg	1505
Qy	840	aggaagtccacagcattggggagataggagttattcagagttcttcacatctctctc	899
Db	1506	aggaagtccacagcattggggagataggagttattcagagttcttcacatctctctc	1565
Qy	900	ggaaacttgaattcaatcagccaccacgctggagaacggttcaaa-----	940
Db	1566	ggaaacttgaattcaatcagccaccacgctggagaacggttcaaaactttggaagacagccgaga	1625

[illegible]

Db 2706 AAGCCTGGTGTCTGCCCCACAGTCCACCAAAAGCTGTAAACAACAACACTCTCTCTTGCC 2765
 QY 1800 tcaactcttaacagagtcctccacagagcaaaccttccagaccacaagcaactagc 1859
 Db 2766 TCCATCCTTTAACAGAGTCTCTCCACAGAGCAAAACCCCTCCAGCCACACAGCAACTAGC 2825
 QY 1860 acaaatgtctccaaacttcaaaactaccattctgaagacctctcagcctcttatgaccac 1919
 Db 2826 ACAAAATGTCTCCAACTTTTCAAACTTCCAACTCTGAAGACCTCTCAGCCTCTTTATGACCAC 2885
 QY 1920 tcttggcacctgtcaagcacagcatctctgttcaactggccttatagccgtacagactac 1979
 Db 2886 TCCTGGCACCTCTCAAGCACACATCTCTGTCTCACTGGCCCTATAGCCGTACAGACTAC 2945
 QY 1980 agtggaaaaacagctctcgtgacccatcctgaaaactagtttccctcaaatctcaacaga 2039
 Db 2946 AGTGTGAAAACAGCTCTCGCTGACCCATCCTGAAATACTAGTTTCTCAAACTCTCAACAGA 3005
 QY 2040 agtggcatcagcacagaaagacagagtgattgtggatgctaccactggattgatccc 2099
 Db 3006 AGTGGCATCAGCACAGAAGAACCGAGTGATTTGGATGCTTACCCTGGATTTGATCCC 3065
 QY 2100 ttbgacagtgatccacatcagcaaaagaaaatgacacaaagcttggggttacagcaga 2159
 Db 3066 TTTGACCACTGTACCCACATCAGCAAAAGAAATGACACAAAGCTTGGCGTTTACAGCAGA 3125
 QY 2160 gtacagccagcttcagcttccctcgaacatctctctcccaaacacacagttgtttc 2219
 Db 3126 GTACGCCAGCTTACAGTTTCCCTCGGAACATCTCCTTCTCCCAACACACAGTTGTTT 3185
 QY 2220 cagggctgaagacttggctcccaaatctgccaccttggcttgcagagcagcacagtc 2279
 Db 3186 CACGGCTGAAGACTTGGCTCCCAAACTGCCACCTTGTGTTTTCAGAGCAGCACAGATC 3245
 QY 2280 accaaacacactgtcctcttcagcctcagtcacacagctgtgtgtgaaccccttgttcca 2339
 Db 3246 ACCAACAACACTGTCTCTTACGCTTACGCTTCAACAGCTGTGTGTGAACCCCTTGTCTTCA 3305
 QY 2340 caatggcgaatcgttcgacagaaacacacagcgtgctaccactgaggtgcccgccttc 2399
 Db 3306 CAATGGCGAATGCGTGCAGACAACACACAGCCGTGCTTACACTGCAAGTGGCCGCTTC 3365
 QY 2400 ctggcaagggtgattgagtgatgtgaatgagtgctgcgaacccctgcccac 2459
 Db 3366 CTGGCAAGGGATGATTGCACTGTGATGTGATGATGCTGTCTGCAACCCCTGCCATC 3425
 QY 2460 cacagccagtgcaacaatactcagggatccttatactgcaaatgcccgttgggtacca 2519
 Db 3426 CACAGCCACGTGCAACAATACTCAGGGATCCTTTATCTGCAAAATGCCCGGTGGGTACCA 3485
 QY 2520 gttggaaaaagggtatgcaatttggtagaaccttcgtgacagagtttaaatgaagag 2579
 Db 3486 GTTGGAAAAAGGATATGCAATTTGGTTAGAACCTTCTGTGACAGAGTTTAAATTAAGAG 3545
 QY 2580 aacttttcttaaataactgtgaaaaacattcagaccctacaagaagtgaatagat 2639
 Db 3546 AACTTTTCTTAATACAACTGTGAAAAAACAATTCAGACCTTACAGAGTTGAAAAATGAGAT 3605
 QY 2640 caccaaaagcttaaatatgtgttttcagcgtttacactagttacatccgatctacagttca 2699
 Db 3606 CACCAAAAGCTTAAATATGTGTTTTCAGCGTTTACCTACTTACATCCGATCTACAGTTCA 3665
 QY 2700 cgccttagggagtcacaagcgtgtgtatctcactgcaaaacaccttttccctggcctc 2759
 Db 3666 CGCCTCTAGGGAGTCCACCGCGTGTGTATCTCACTTGCAAAACAACCTTTTCCCTGGCCTC 3725
 QY 2760 caatgtgacgtatttgacctggctgataggatgcagaaaatgtgtcaactcctgcgaagtc 2819
 Db 3726 CAATGTGACCTATTTGACCTGCTGATAGGATGCAGAAATGTCAACTCTCTGCAAGTC 3785
 QY 2820 ctctgtgaggtctgcagctcttggatctcagagggcgatctttagagcggcgacgtt 2879

Db 3786 CTCTGTGAGGTCTGCGAGCTCTTGGGATCTCAGAGCGGATCTTTAGAGCGGGCAGCTT 3845
 QY 2880 gtcaagcggaaagagtccegaatagtgacaaagacacacctccatctgcaactgaacctggaagg 2939
 Db 3846 GTCAAGCGGAAGAGTCCCGAATGTGACAAAGACACCTCTCCTGCACTGACCTGGACGG 3905
 QY 2940 cgttgcctctgcagtgcaagtgcgggatacttcaagtccaagaagtgaacacactcctg 2999
 Db 3906 CGTTGGCCTGTGCCAGTGCAGTCAAGTCCGGATACTTTTCAGTTCAACAAGATGACCACTCTCTG 3965
 QY 3000 ccgagcatgtgaagatggatataggtctgaaatgaacactgcatgagttgcccatttgg 3059
 Db 3966 CCGAGCATGTGAAGATGGATATAGGCTTGAATGAAACCTGCAATGAGTTGCCATTTGG 4025
 QY 3060 ccttgggtgctcaactgtgaaaaacccctatacagcttatcactgtgtgtatcgacgcgc 3119
 Db 4026 CTTTGTGTGTCTCAACTGTGAAACCCCTATCAGCTTATCAGCTGTGTGTGATCGCAGCCGC 4085
 QY 3120 gggaggtgggtcctcctcctccttaggcacgcactgattgttacctgttgcgagaagaa 3179
 Db 4086 GGGAGGTGGGCTCTCTCATCTAGCATCGCACTGATTTTACCTGTTTGCAGAAAGAA 4145
 QY 3180 taaaaatgacataagaacaaactccttcaaaagtggagatttccaaaatgtccccatagc 3239
 Db 4146 TAAAAATGACATGAACAAACTCATCTTCAAAAGTGGAGATTTCCAAATGTCCCCATATGC 4205
 QY 3240 tgaataccctcaaaaaactcctcgtccacaagaatggggccgagaagctattgaaatgcata 3299
 Db 4206 TGAATACCCCAAAATTCCTCGCTCACAAAGATGGGGCCGAGAGCTATTGAAATGCATGA 4265
 QY 3300 gaatggaagtaccaaaaacactcctccagatgaaggatgtgtactactcgcctcacaagtgt 3359
 Db 4266 GAATGGAAGTACCAAAAAACCTCTCCAGATGACGGATGTGTACTACTCGCCTACAAGTGT 4325
 QY 3360 aaggaatccagaactgaaacgaacggactctaccogggctacactggaactgcccagatc 3419
 Db 4326 AAGGAATCCAGAACTTGAACGAACAGGACTCTACCCGGCTTACACTGGACTGCCAGGATC 4385
 QY 3420 acggactctctgcatcttcccgagacagataacccgtcttctcatcagtgatgaagacag 3479
 Db 4386 ACGGCATCTTGTGCAATTTCCCGGACAGTATTAACCGTCTTTTATCATGATGATGAAACGAG 4445
 QY 3480 aagaagagactacttttaagtcacaggagagagaggactcatctgctgagccagtcacc 3539
 Db 4446 AAGAAGAGACTACTTTTAAGTCCAGGAGAGAGAGGGACTCATTTGCTGTGAGCCAGTCACC 4505
 QY 3540 tgggaactcgtctcagagagccacacagagactgcccagagatttgcggagaccac 3599
 Db 4506 TGGGACCTCTGCTCAGAGGACCCACAGAGGGCTGGCCCGCAGGATTTGTTCGGGAGCCAC 4565
 QY 3600 gctgagtggaagcaggaagaggagacagagcagtcgcgggcgtgacacagtggaaggagaca 3659
 Db 4566 GCTGAGTGGCAACGACGAGAGGGACAGGATCGGGGCGTGACCAAGTGGAGGAGACA 4625
 QY 3660 ggtggatgtggaaccacaggtgctcactcaagcaccttctgttactctggaacgtgaat 3719
 Db 4626 GGTGGATGTGGAAACCAACAGGCTGCTATTTCAGCACCTTTTGTGTACTGTGAACGTGAAT 4685
 QY 3720 gttgggacagtatcaagagactctctgagtgactgacacactgacacccagtcggacccggcg 3779
 Db 4686 GTGGGCGAGTATCAAGAGAGTCTCTCTGAGTGACTGCAACCTGGCACTGGCAGCCAGGGCG 4745
 QY 3780 actatlagccagggcagaccactagactcctcagtgaggagacctgtttccctcctgttgg 3839
 Db 4746 ACTATTAGCAGGCGACACCACTAGACTTTCAGTGCAGGGACCTGGTTTCCCTTCGTTTG 4805
 QY 3840 cacttttagtaaatgtgggtggaggttcccttttggatctcttbtggactgttttccagaaa 3899
 Db 4806 CACTTTAGTAATAATGGGTGGGAGTTTCCCTTTTGGATCTCTTTTGTGACACTGTTTCCAGAAA 4865
 QY 3900 gaaggtctcttcccgagagacacttccatagcagcaaatgtgtgatttcttcttgcagcaa 3959
 Db 4866 GAAGGCTTCCCTTCCCGAGACACTTCCATAGGCAGCAAAATTTGGTGATTCATTTGTCASCA 4925

QY 3960 aactagctgttgaattattttctccagcgcctgcgtctaaacacagatgagga 4019
DB 4926 AATACTGGCTGTGTAATATTTTCTGCCAGCACCCTGCGTGTAAACACAGATGAGGA 4985
QY 4020 tgaagctaccactgaagctgaagatgcctcattgaacgagacagtgcttttcatatgttt 4079
DB 4986 TGASCGTACCACCTGAAGTCTGAAGATGCGCAATGAACGGACAGTGTGTTTCATATGTTT 5045
QY 4080 ctgaagtgcttattgctacagtttccaaagccagcccccacagtgaggaatgtgtgaggg 4139
DB 5046 CTAGGTGTCTTATGCTACAGTTTCCAAAGCCACSCCCACAGTGAGGAATGTGTGAGGC 5105
QY 4140 accgcacacactcaatgtgtttttaaagtcagggtgacacatgtatttaagatttttt 4199
DB 5106 ACCGCACACAACTGCATGTGTTTAAAGTCAAGGTGACACATGTATTTAAGATTTTTT 5165
QY 4200 tttaaaatctcttgcagtttaaatctcaacttttttcaaacagcctggatcagggcgaaac 4259
DB 5166 TTTAAAAATCTCTTGCAGTTTAAATCTCACTTCTTCAAAACAGCCTGGATCAGGGCAAAAC 5225
QY 4260 aacttatatttggttagctgaggtcagcagcagcagcagcagcagcagcagcagcagcagc 4319
DB 5226 AACTATATATGTGTTTAGCTGAGGCTCAGCAGGCGAGATTCAGGCGAGGGGGGACACTTT 5285
QY 4320 tcatccatgagggccagcctgggacctgggacctctgatcaccattgtgagggccaggg 4379
DB 5286 TCATCCATGAGGGCCAGCCTGGGGCCTGGGACTCTGATCACCATTGTGAGGGCCACAGG 5345
QY 4380 cagctgcgtatggaggagaaatgtcaaacctgaacgcaggtttccaccactcttagaaagca 4439
DB 5346 CAMCTGCGTATGGAGGAGAAATGTCAAACTGAACGCGAGGTTTCCACCCTCTAGGAAAGCA 5405
QY 4440 gcttgtagccctcagctgagtgatgtggttagagggatgggctgaatagcaggttaga 4499
DB 5406 GCTTGTGTGACCCCTGCACCTGGATGTGGTGTAGAGGATGGCTGAATAGCAGGTTTGA 5465
QY 4500 ttctctcatcaacagctgttgggaagctgtgtggatctcctgaggaagacagggagcc 4559
DB 5466 TTTCTCTCATCAACAGTGTCTTGGGAASCTGTGTGGATTCTTGAGGAAGAACAGGGAGCC 5525
QY 4560 gagatgagccacacatgagttgtctcaccggctactgcagcactttgtaccacagaatct 4619
DB 5526 GAGATGGAGCCACACATGAATTTGCTCACCCTGCTACTGCAGCACTTTGTACCCAGAACT 5585
QY 4620 catgtcaaaaacccatgtaaactttcaaccactcaaaactgtttattcggctgaagaa 4679
DB 5586 CATGTCCACAAACCCCATGTAAACCTTTCAACCCACTCAAAAGSTGTTTATTCGGCTGAAGAA 5645
QY 4680 ataacttttttctcaccagtcattgttacctcttcacatgctgctgtgcaccctcc 4739
DB 5646 ATAACTTTTCTTCTACCCAGTCACTTTGTACCTTCTCATATGGSATATGCGCACCCCTCC 5705
QY 4740 agaaactggtattacttccagtcagtggtgggaactgaagacttccggttggctcaggg 4799
DB 5706 AGAAACGTGGTTATATKCCAGTCCAGTGTGGGAGAACTTCCCGTGTGTCAGG 5765
QY 4800 aactgaaggttgactcgggaaggaagtgatccactcatttatttattatgctctgagtg 4859
DB 5766 AACTGAGGGTGTACCTTCGGGAAGGAAGTTCCACTCATCTATTTATTTATGCTGTGATG 5825
QY 4860 tgggtctcgcaggagacatcagtcagtgctcttcaattgcccactcgggaaactgt 4919
DB 5826 TGGGTCTCTGCAGGGAGACATCCAGTACTCGGTCTCTKATTTGCCACCTGGGGAACGTG 5885
QY 4920 gtttattggcctcttttggggcatcctgggttttggatgaagtgagggggaatacagaggtta 4979
DB 5886 GTTTATTGGCTCTTTTGGGCATCCTTGKTTCCGATGAAAGTGAAGGGAATACAGAGGTA 5945
QY 4980 aaagaatgtctcaacctgaagcggggagtcocgcttcacatttctggaaatggtgcag 5039
DB 5946 AAAGAATGTCTCACCCCTGAAGCGGGAGTCCCGCTTCACATTTCTGGAATGTTGTCAG 6005

QY 5040 ccactgggacagttctgccccggcagtcagttgttttcaaggtcctctaaataaatac 5099
DB 6006 CCACTGGGACAGTCTGCCCCGGGCATGGTGTGTTCTTCAAGGTCTCTTAATATAATC 6065
QY 5100 cctattttacataatccttggccctgatggttttaagcaagaactcctgtgtcccatgg 5159
DB 6066 CCTATTCTTACATAATCCTTGGCCCTGATGTTTTTAAGCAAGAACTCCTGTGTCCMATGG 6125
QY 5160 tctccacactcaacatacaccctctgtagcaagagtccttagtcaggggaggtgcatttt 5219
DB 6126 TCTCCACACTCACCATCACCCCTGCTGTAGCAAGAGTCTTAGTCAGGGGAGGTGCATTTT 6185
QY 5220 agtagttaaattgacacttatccatgagataaataaaagaggaactgttttlatcagtgga 5279
DB 6186 AGTAGTTACATTGTCATTCATGAGATAAATAAAGGAGAVCTGTTTTTATCAGTGA 6245
QY 5280 ggctaaactaaaattcaaaagtgtcgcccttttgaaatcttgggcccctctctctgtaga 5339
DB 6246 GGCCTAACCTAAATTTCAAGTGTGCGCTTTTGAATCTTTGGGCTCTCTCTCTGTAGA 6305
QY 5340 accaatggcccttggctcagcagcctcgcacactaactgagagattctgagctcctgca 5399
DB 6306 ACCAATGCCCTTTGTGGCTCACGGCCTCGCACCTAACTGGAGAGTTCTGAGCTCCTGCA 6365
QY 5400 gctcactgagccacacagactaggtcttctggtcctctccgc 5441
DB 6366 GCTCACTGAGCCACACAGACTAGGCTTCTTGGCTCTCTTCCGC 6407

RESULT 5
US-08-944-496-7
: Sequence 7, Application US/08944496
: Patent No. 6124433
: GENERAL INFORMATION:
: APPLICANT: FALLB, DEAN A
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
: TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
: NUMBER OF SEQUENCES: 54
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/944,496
: FILING DATE: 06-OCT-1997
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/599,654
: FILING DATE: 09-FEB-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/485,573
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/386,844
: FILING DATE: 10-FEB-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: CORUZZI, LAURA A
: REGISTRATION NUMBER: 30,742
: REFERENCE/DOCKET NUMBER: 7853-104
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:

```

; LENGTH: 6407 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
US-08-944-496-7

```

Query Match 73.3%; Score 5098.4; DB 3; Length 6407;
Best Local Similarity 94.3%; Pred. NO. 0;
Matches 5412; Conservative 21; Mismatches 8; Indels 301; Gaps 2;

Qy	1	gttcgatgaagaattgcgcgcttttcaaacaaagagtggaaacagcctcgagatgggaac	60
Db	666	GTTTCGATGAAGAATAATCCCGCTTTTCAAACAAGAGTGCAACAGCCTCGAGATGGGAAC	725
Qy	61	agagaggcgatgggctatcagaataaggactgtcacagccaagagggccacaacttc	120
Db	726	AGAGAGGCCGATGGGCTGTCAAGAANTGGACTGTGCACACGAAGAGGCCACCACATTTC	785
Qy	121	ggcttgagagcccttccttctctctcttggagatgggagagctgacacgccctcttag	180
Db	786	GGCTTGGAGCCGCTCTTCTTCTTCTGCTTTGGAGATGGGAGAGCTGACCACGCCTTCTAG	845
Qy	181	gaagagaaaattcttcaggaccagatctctctctgctgctgaattctctacagagacagcttc	240
Db	846	GAAGAGAAATTCCTCAGGACCAGATCTCTCCTGGTGCATTTCTACAGACAGCAGCTTC	905
Qy	241	ctctctctcttagaccttctctcaacctctgaagtacagagaagcttaacaactccac	300
Db	906	CTCTCTCTCTTAGACCTTCTCTCACCTTCTGAAGTACAGAGAAGCTTAACAGACTCCAC	965
Qy	301	tggcctccagagctctctcagtcagtcacaaacaagacaaatgcattgttgtaccggtttcac	360
Db	966	TGSCCTCCAGAGCTCTCAGTCAGTCAACAAAGACAATGCATGTTGCTACCGTGTTCAC	1025
Qy	361	tgatgtgcccagagaagcgtcgatctcttgacgcgtcagtcgggacctgtgagcaagac	420
Db	1026	TGATGTGGCCCCAGAACGCTCGAATCTTTGACGGTCTAGTCTGGGACCTGTGAGCAAGAC	1085
Qy	421	agaagcttccccagaagactccagaattgccagacttcatcctcagtccttcttctacc	480
Db	1086	AGAAGGCTTCCCCAGAGCTCCAGAAATGCCACGACTTCATCCTCAGTCTCTTCTTTCACC	1145
Qy	481	ctctgcagtggaatcgagagaagaaacagtagatgaactgsggaatccaggggatgaggaatt	540
Db	1146	CTCTGCAGTGGNAATCAGAGAACAACAGTAGATTAAGTGGNATCCAGGGGATGAGCAATT	1205
Qy	541	cattgaacctccacagaaaaatgaatttggactacgtcttcttcg - tggcaaaaatgattc	599
Db	1206	CATTGAACCATCCACAGAAAATGAATTTGGACTTACGCTCTTTGGCTTGCAAAAATGATTC	1265
Qy	600	cccaaccttggagaacatacagcttgcacagagctctcaggtgcacaaatggaagtcccat	659
Db	1266	CCCAACCTTTGGAGAACATACAGCTTGCCACAGCTCTGAGGTGCAAAAATGGAAGTCCCAT	1325
Qy	660	gtctcagactgagactgtcttagtctcagtcgcacccatgagagctgagagatcaactgc	719
Db	1326	GTCTCAGACTGACACTGTCTTAGTTCAGTCCGCCACCCATGAGAGGTGGAGAGATCACTTCG	1385
Qy	720	acactggctcttgaccaacagcacacaactctgcagatgtgacaggaagctctgtctcata	779
Db	1386	ACACTGGCTCTTGACCAACAGCACACAATCTGCAGATGTGACAGGAAGCTCTGCTTCATA	1445
Qy	780	tccatgaaggtgtgaatgccttcagvtgtgaccacagttctccagactactgtacagctctgg	839
Db	1446	TCCCTGAAGGTGTGAATGCTTTCAGTGTGACCCAGATTCTCAGACTCTACTGTACAGTCTGG	1505
Qy	840	agaaatcacacagacttggagatagaggttatccagagcttctcatcatctctctc	899
Db	1506	AGGAAGTCAACAGCAATGGGAGATAGGAGTATTTCAGAGTCTTCATCTACATCTTCCTC	1565
Qy	900	qgaaagcttgaattcatcatgaccacacagtcagtgagaaacgtttcaa -----	940

Db	1566	GGAAAGCTTGAAATTCATCATCAGCACCGCTGGAGAACGTTCAACCTTTGGAAGACACGCCGAGA	1625
Qy	941	-----	940
Db	1626	GCCAGGCCAAGCACTAGGTGCACAGTTCCGCCCAATGCAGAGGACAGAGACTTCTGGGGTGCC	1685
Qy	941	-----	940
Db	1686	CTCTCTCGGCACCCACACACCTTGCTGTACTGTCTACTGGAAACGGGGAACGCACACTCGCGTC	1745
Qy	941	-----	940
Db	1746	TGTCACCCCTCACCACACACAGCATGAGCAGAGACTTCTGGGGAAGCAGCGCCCTGCAGC	1805
Qy	941	-----	940
Db	1806	GGCCATGCCCAAGAAACAGAGGGTGCCCTCTCTGCACGTTAAACGTGAACGTCAGCGAGCATGGG	1865
Qy	941	-----tcgcctgggattagctacgg	959
Db	1866	CCTGGTCTCAGGTCACCTGGCCGCTCCAGTCGCCTGGGAGTCGGCTGGGATAGTATCAGG	1925
Qy	960	tcaagtcgtggcacagctattgaacaaaggacttccagcgaccacacagaccacaccta	1019
Db	1926	TCAAGTCGCTGGCACAGCTATTGAACAAAGGACTTCCAGCGACCACACAGACCACACCTA	1985
Qy	1020	ccgtcctatctacitttcaccaaaaggaaacggcggttactgttcattacagataaacagttc	1079
Db	1986	CCTGTCATCTACTTTACCACAAAGGAACGGGCGTGTCTGTCATTCACAGATAACAGATTC	2045
Qy	1080	atcctcagacattgtggagagctccaactctctatatataaaatctcaaaactcttcacattc	1139
Db	2046	ATCCTCAGACATTGTGGAGAGCTCAACTTCTATTATTAATCTCAAACTCTTCACATTC	2105
Qy	1140	agadcttcctctcttttctcatgtctcagactgagagaagtaacatctcatcctatgaagg	1199
Db	2106	AGAGTATTCCTCTCTTTCTCATGCTCAGACTCAGAGAGAAGTAACATCTCATCTATGACGG	2165
Qy	1200	ggaatagtcagccctctactagtcgcgcagttctgtacatcacatccaaacctccgctcta	1259
Db	2166	GGATATGCTCAGCCTTCTACTGAGTCGCCAGTTCTGCATACATCCACCTTCGTCCTA	2225
Qy	1260	cacccccaccattaaatgcccgaacacttcggtgtgtctctggacactgatgtgtgatttgt	1319
Db	2226	CACACCCACCATTAATATGCGGAACACTTCGGTGTGTTCTGGACACTGATGTGTGATTTGT	2285
Qy	1320	tagtgactcaggccctctt	1379
Db	2286	TAGTGACT	2345
Qy	1380	gcctctgccctctgtgtccaaatcccaacattattttccaatctttaccatcaaacag	1439
Db	2346	GCCCTCGCCCTCTGTGTCAATATCCACACATTAATTTTCATCAATTTTACCATCAACAG	2405
Qy	1440	ggcctctgtgcatctactaaagtctactctctgtatgcataccacacacacacacacacacac	1499
Db	2406	GGCCTGTGCATCTACTAAAGTCTACCTTCTGATGCATCCACACCATGGTCTCTCTCACC	2465
Qy	1500	atcacctttaccagtagtctcttaacgacatctacatctgtcccaactttctgtctcacaaac	1559
Db	2466	ATCACCTTTACAGATATCTCTTAACGACATCTACATCTGCCCCACTTCTGTCTCACAAAC	2525
Qy	1560	aaccttgccacagtcatcttctacccctctctgtcccgagggaaggagactcctgtgac	1619
Db	2526	AACCTTGCCACAGTCATCTTCTACCCCTCTCTGCCAGGGGAGAGACTCTCTGTGAC	2585
Qy	1620	ttaatttcagaactaacatgacatcatctatgacaatgtccatagtaactcaactgc	1679
Db	2586	TTCAATTTACAGACATCAACATGACATCATTTCAATGACAATGCTTCCATAGTGTCAAACTGC	2645
Qy	1680	agaccttaagagccagagcaccaccaaccaagagaagtcattacagaaatcaaaagtcaacc	1739

Db 2646 AGACCTTAAGAGCGCAGAGCACCCACACCAAGAGAAAGTCATTTACAGAATCAAAAGTCACC 2705
QY 1740 aagcctggtgtctgtccacagagtcaccaaagctgtacaaacaaactctcttcttgcc 1799
Db 2706 ANGCTGGTGTCTGTGCCACAGAGTCCACAAAGCTGTAAACAAACACTCTCCTTTGGCC 2765
QY 1800 tccatctctaacagagtcctccacagagcaaaccttccagccacaaagcacaacttagc 1859
Db 2766 TCCATCTTAACAGAGTCTCCACAGAGCAAAACCTTCCAGCCACAAACCAACTTAGC 2825
QY 1860 acaaatgtctccaactttcaacactaccattctgaaagacotctcagcctcttatgaccac 1919
Db 2826 ACAAAATGCTCCAACTTTCAACACTACCAATCTGAAGACCTCTCAGCCTCTTATGACCAC 2885
QY 1920 tccctggcacctgtccaagcacagcatctctggttcacttgccctctatagccgtacagactac 1979
Db 2886 TCCCTGGCACCTGTGTAAGCACAGCATCTCTGGTCACTGGGCCCTATAGCCGTACAGACTAC 2945
QY 1980 agctggaaaaacagctctcgttgacccatcctgaaataactagtctcctcaaatctcaacaga 2039
Db 2946 AGCTGGAAAACAGCTCTCGCTGACCCATCTCTGAANAATCTAGTTCTCTCAAAATCTCAACAGA 3005
QY 2040 aggtggcatcagcacagaaaggaaacaggatgattggtggtctaccactgattgatccc 2099
Db 3006 AGGTGGCATCAGCACAGAAAGGAACCCAGTGTATTGTGGATGCTTACCACCTGGATTGATCCC 3065
QY 2100 ttctgaccagtgtaaccacatcagcaaaagaataagcacaaaagcttggttaccagaga 2159
Db 3066 TTTGACCAGTGTACCCACATCAGCAAAAGAAATGACCACAAAGCTTGGCGCTTACAGACAGA 3125
QY 2160 gtacagcccagcttcagcttccctcggaacatctccttctcccaaacaccagctgtttc 2219
Db 3126 GTACAGCCAGCTTCAGTTCCCTTCGGAACATCTCCCTTCTCCCCAAACACAGTTGTTTC 3185
QY 2220 cacggtggaagacttggctcccaaatctgcacacttggcttctcagcagcagcacagtc 2279
Db 3186 CAGGCTGAAGACTTGGCTCCCAAAATCGCCACCTTGGCTGTTTCAGAGCAGCACAGTC 3245
QY 2280 accaaacacactgtctcttcagctcagtcacacagctgtgctgtgaaccttgtcttca 2339
Db 3246 ACCAAACACACTGTCTTCAGCCTCAGTCAACAGCTGTGCTGTGAACCTTGTCTTCA 3305
QY 2340 caatggcgaatgctgcagacaacacagccgtgggtaccactgcaggtgcgcgccttc 2399
Db 3306 CAATGGCGAATGCTGCAGACAAACACAGCGGTGGCTACCAGTGCAGGTGCCCGCTTC 3365
QY 2400 ctggcgaagggaatgctgcagtggtggtggaatgagtgctgtcgaaacccctgcccac 2459
Db 3366 CTGGCAAGGGGATGATTGCAAGTGTGGATGTGAATGAGTGGCTGTGCGAAACCCCTGCCCATC 3425
QY 2460 cacagccagtgcaacaataactcaggatccttattctgcaaatgccggttgggtacca 2519
Db 3426 CACAGCCACGTGCACAAATACTCAGGGATCCTTTATCTGCAAAATGCCCGTTGGGTACCA 3485
QY 2520 gttggaaaaaggatatgcaatttggtagaaccttctgtacagagtttaaatataagag 2579
Db 3486 GTTGGAAAAAGGATATGCAATTTGGTTAGAACCTTCGTGCACAGAGTTTAAATTAAAGAG 3545
QY 2580 aactttcttaatacactdggaaaaacattcagacctcaagaagtgtgaaatgagat 2639
Db 3546 AACTTTTCTTAATACAACTGTGGAAMAAACATTCAGACCTACAAGAAGTTGAAAAATGAGAT 3605
QY 2640 caccaaaacttaaatatgtttttcagcgttaccttagttacatccgatctacagttaa 2699
Db 3606 CACCAAAACCGTTAAATATGTGTTTTTCAGCGTTACCTTAGTTACATCCGATCTACAGTTCA 3665
QY 2700 cgccctagggagtcacaacgggtgggtgatatctcactgcaacaacacctttccctggcctc 2759
Db 3666 CGCCTTAGGGAGTCCAAACGGGTGGTGATCTCACTGCAACAAACCTTTTCCCTGGCCTC 3725
QY 2760 caatgtacgctatttgacctggctgataggaatgcagaaatgtgtcaactcctgcgaagtc 2819
Db 3726 CAATGTGACGCTATTGTACCTGTGATAGGATGCAGAAATGTGTCAACTCCTGCAAGTC 3785

QY 2820 ctctgctgaggtctgcagactcttgggactctcagaggcgagatcttttagagcggcagctt 2879
Db 3786 CTCTGCTGAGGTCTGCAGGCTCTTTGGGATCTCAGAGCGGATCTTTAGAGCGGCAGGCTT 3845
QY 2880 gtgcaagcggaagagtcctccgaatgtgcaaaagacacctccatctgcactgcactgcactgcacgg 2939
Db 3846 GTGCAAGCGGAAGAGTCCCGAATGTGACAAAGACACCTCCATCTGCACGTGACCTGGACGG 3905
QY 2940 cgttgccctgtgcagtgcaagtcgggatactttcagttcaacaagatggaccactcctg 2999
Db 3906 CGTTGCCCTGTGCCAGTGCAGTCCGGATACTTTTCAGTTCAACAAGATGGACCACTCCTG 3965
QY 3000 ccgagcaatgtgaagatggaatagagcttgaaaaatgaaacctgcagtgtgagttgcccatttgg 3059
Db 3966 CCGAGCATGTGAAGATGGATATAGGCTTGAANAATGAAACCTGTCATGTGTCCTCATTTGG 4025
QY 3060 ccttgggtggtcctcaactgtggaacccctatcagcttctcactgtgtgagtcgcagccgc 3119
Db 4026 CTTTGGTGGTCTCAACTGTGGAAACCCCTATCAGCTTATCACTGTGTGTGATCGCAGCGCG 4085
QY 3120 gggagggtggctcctgctcatcctaggcatcgcactgattgttaccctgtgtgcagaaagaa 3179
Db 4086 GGGAGGTGGGCTCCTGCTCATCTCCTAGGCATCGCACCTGATTGTTACCTGTTGCAGAAAGAA 4145
QY 3180 taaaaatgacataaagcaaaactcctctcaaaagtggagatttccaaatgtcccccatatgc 3239
Db 4146 TAAAAATGACATAAGCAAACTCATCTTCAAAGTGGAGATTTCCAAATGTGCCCATATGC 4205
QY 3240 tgaatcccccaaaaactcctcgtcacagaatggggccgagaaactattgaaaaatgcata 3299
Db 4206 TGAATACCCCAAAAANTCCTCGCTCAAGAATGGGGCCGAGAGACTATTGGAATGTCATGA 4265
QY 3300 gaatgggaagtaccaaaacccctccagatgaecggaatgtgtactactcgcctcaaatgt 3359
Db 4266 GAATGGAAGTACCAAAAACCTCCTCCAGATGACGGATGTGTACTACTCGCCTACAAAGTGT 4325
QY 3360 aaggaaatccagaaacttgaaagaaaaggactctacccggcctacactggactgcaggatc 3419
Db 4326 AAGGAATCCAGAACTTGAACGAAACCGGACTCTACCCGGCTTACACTGGACTGCGCAGGATC 4385
QY 3420 acggcatcttctgatttccccggacagataaaacccgtcttctcatcagtgatgaagcag 3479
Db 4386 ACGCATCTTGCATTTTCCCGGACAGTATAACCCGTCTTTTCATCAGTGTGATGAAGCAG 4445
QY 3480 aagaagactacttttaagtccaggagagagggactcaatgtctctgagccagtcacc 3539
Db 4446 AAGAAGAGACTACTTTTAACTCCAGAGAGAGAGGGACTCATTTGCTCTGTAGCCAGTCAAC 4505
QY 3540 tgggacctctgctcagagagaccgaccagagagctgcgccaggatttgcgggagccac 3599
Db 4506 TGGGACCTCTGCTCAGAGGACCCGACCAGGAGGTGGCCCGCAGGATTTGTGGGAGGCCAC 4565
QY 3600 gctgagtggaagcaggaagggagcagggcagtcgvgggcgtgaccacagtgagggagaca 3659
Db 4566 GCTGAGTGGCAACAGCAGAGGGGACAGCATCGGGCGTGCACCAAGTGGAGGAGACA 4625
QY 3660 ggtggaatgtggaaccacagagctgctcatcagcacccttctgttactgtgaacgtgaat 3719
Db 4626 GGTGGATGTGGAAACACACAGGCTGCTCATTTACGACACCTTTGTTGTTACTGTGAACGTGAAT 4685
QY 3720 gtagggccagtatcaagagagctctctgagtgactgcacatgagcactgcacccagggcg 3779
Db 4686 GTGGCCAGTATCAAGAGAGTCTCTGTGAGTGTGACCATGTCATGTCACCGAGGCGG 4745
QY 3780 actattagccaggcgacagccactagacttcagtcaggggacctgtgtttccctctcgcttg 3839
Db 4746 ACTATTAGCCAGGCACACACCTAGACTTCAGTGCAGGGACCTGGTTTCCCTTCGTTCG 4805
QY 3840 cactttagtaaatgtgggtggaggttccctttggatctctgttttagactgttccagaaa 3899
Db 4806 CACTTTAGTAATTTGGGTGGAGGTTTCCCTTTTGGATCTGTGTTTGTGAGACTGTTCCAGAAA 4865

Qy 3900 gaaggcttccttcccgagacacttccatagggcagcaatttgggtgattcatttgcagcaa 3959
Db 4866 GAAGGCTTCCYTTCCGAGACACTTCCATAGGCAGCAATTTGGTGATTCATTCASCAA 4925
Qy 3960 aactagctgttaattatttcttcccgagcgcctgctgctaaacaacagatgagga 4019
Db 4926 ANTACTGGCTTGTAAATATTTTCTGCCACAGRCCTGCGTGTAAACAACAGATGAGGA 4985
Qy 4020 tgagctaccactgaagctgaagatgtcgccattgaacgacagtggtttccatagtgtt 4079
Db 4986 TGASCGTACCACCTGAAGTCTGAAGATGTGCCATTGAACGACAGTGTTCATATGTTT 5045
Qy 4080 ctaggctgtcttatgtcacagtttccaaagcccccacagtgaggaatgtgtgagcc 4139
Db 5046 CTAGGTTGTCTTATGCTACAGTTCCTCAAGCCASCCCCACAGTCAGGAAATGTGTAGGC 5105
Qy 4140 accgcacacactcaatgtgtttttaaagtcgaagtgacacatgtatttaagattttt 4199
Db 5106 ACCGCACACACTGCATGTGTGTTTAAAGTCAAGGTGACACATGTATTAAAGATTTTTT 5165
Qy 4200 tttaaaatctcttgcagtgtaaatctcaccttttccaaacagccctggatcagggcaaac 4259
Db 5166 TTTAAATCTCTYTGCAAGTTAAATCTCACCTTTTCAAAACAAGCCTGGATCAGGGCAAAAC 5225
Qy 4260 aactatatttggttttagctggaagctcagcagcgagtgagcagggggcacttt 4319
Db 5226 AACTTATATYTGTTTAACTGGAGGCTCAGCAGGCAGATTGCAAGCAGGGGGGCACTTT 5285
Qy 4320 tcaccatgaggcccgagcctgggagctggagactcgtacacatgttgagggccagagg 4379
Db 5286 TCATCCATGAGGGCCAGCCTGGGGCTTGGGACTCTGATCACCATTGTGAGGGCCAGAGG 5345
Qy 4380 cagctcgatggaggagaaatgtcaaacactgaacgaggtttccaccactctaggaagca 4439
Db 5346 CAMCTGCTATGGAGGAGAAATGTCAAACTGAACGCAGGTTTCACCACTCTAGGAAGCA 5405
Qy 4440 gcttgtagccctcgcagctgtagtgctggttagagggatgggctgaatagcaggttaga 4499
Db 5406 GCTTGTGACCCCTCGCASCCTGGATGTGTTAGAGGGATGGCTGAATAGSCAGGTTAGA 5465
Qy 4500 ttctctgcaaacagtgcttgggaagctgtgtggattctcaggaagcagggagcc 4559
Db 5466 TTTCTCTGCATCAACAGTGTCTTGGGAASCTGTGTGGATTCTGAGGAGAACAGGGAGCC 5525
Qy 4560 gagatggagccacacatgagttgtctcaccggctactgcagcactttgtaccagaaact 4619
Db 5526 GAGATGAGCCACACATGAATTGCTCACCCTGCTACTGCAGCACTTTGTACCCAGATCT 5585
Qy 4620 catgtccacaaacccatgtaaactttccaaaccactcaaaactgtttatcggtgaagaa 4679
Db 5586 CATGTCCACAAACCCCATGTAAACTTTCAACCACTCAAAAGSTGTTTATTCGGCTGAAGAA 5645
Qy 4680 ataacttttttctaccagcattgttaacctttcatatggctgtgcaaccctcc 4739
Db 5646 ATAACTTTTCTCACCAGCATTTGTACCTCTTCATATGSGTATGTCGACCCCTCC 5705
Qy 4740 agaaactgtgtatacttccagtcagtggtgggagaactgaagcttcggttggtgag 4799
Db 5706 AGAACTGTGTTATCTKCCAGTCAGTCTGGGAGAACTGAAGACTTCGGTTGTCAGG 5765
Qy 4800 aactgaggggtgaccttcgggaaggaagttccactcatcttattattatgcctgtgatg 4859
Db 5766 AACTGAGGGTTGACCTTCGGGAAGGAAGTTCCACTCATCTTATTATTATGCTGTGATG 5825
Qy 4860 tgggtctgcagggagacatccagtaactcgggtgtcttttaattgcaacctggggaactgt 4919
Db 5826 TGGGTCTGCGAGGGAGACATCAGTACTCGGTGTCTTKTAATTGCCACCTGGGGAACTGT 5985
Qy 4920 gtttattggcctctcttggggcatcctgggttttggatgaagtgagggaatacagaggtta 4979
Db 5886 GTTTATTGGCTTCTTTGGGGCATCTTGKTTGGATGAAGTCAAGGGAATACAGAGGTA 5945
Qy 4980 aaagaattgtctccaaacctgaagcgggagtcctcgtttcacatttctggaaatggtgcag 5039

Db 5946 AAAGAATTTGTCTCACCCCTGAAGCGGGAGTCCCGCTTCACATTTCTGAAATGGTGCA 6005
Qy 5040 ccaactggggacagttctgccccgggcatggtgtgtttcttcaaggtcctcctaaataatc 5099
Db 6006 CCACCTGGGGACAGTTCTGCCCGGGCATGTTGTTCTTCAAGGTCCTCTAAATAATC 6065
Qy 5100 cctattcttacaataatccttgccctgatgggttttaagcaagaactcctgtgcccag 5159
Db 6066 CTTATTCTTACATTAATCTTGCCCTGATGTTTAAAGCAAGAACTCCTGTGTCMATGG 6125
Qy 5160 tctccaaactcaccatcaacctcgtgtagcaagtgctcctagtcaggggaggtgcatttt 5219
Db 6126 TCTCCACCACTTCCATCACCTGCTGTAGCAAGAGTCTTAGTCAGGGGAGGTGCATTTT 6185
Qy 5220 agtagtttaaatgcacttatccatgagataaataaaaggagaactgttttatcagtga 5279
Db 6186 AGTAGTTACATTTGACATTATCCATGAGATAAATAAAGGAGAVCTGTTTTATCAGTGA 6245
Qy 5280 ggctaaacctaaatttcaaaagtgcgcctttttgaaactctgggctcctctctctgtaga 5339
Db 6246 GGCTAACTTAAATTTCAAAGTGTGCTTTTGAATCTTTGGGCTCTCTCTCTGTAGA 6305
Qy 5340 accaatggcccttctgctgctcagcggcctcgcacctaactgaagagttctgagctcctga 5399
Db 6306 ACCAATGCCCTTTGTGGCTCAGCGCCTCGCACCTTAACCTGAGAGAGTTCTGAGCTCCTGCA 6365
Qy 5400 gctcacctgagccacacagactagcttcttgctcctctccctccgc 5441
Db 6366 GCTCACCTGAGCCACAGACTAGGCTTCTTTGGCTCCTTTCCGC 6407

RESULT 6
US-09-328-111-83/c
; Sequence 83, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 83
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(584)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-83

Query Match 6.8%; Score 470.4; DB 4; Length 584;
Best Local Similarity 95.9%; Pred. No. 4.6e-121;
Matches 544; Conservative 0; Mismatches 16; Indels 7; Gaps 6;


```
; LENGTH: 128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; US-08-485-573-7

Query Match          1.8%; Score 126.4; DB 2; Length 128;
Best Local Similarity 99.2%; Pred. No. 1.2e-25;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3120 gggaggtgggctctctatccttaggcacgcactgattgttacctgttgcagaaagaa 3179
Db 1 GGGAGGTGGGCTCTCTGCTCATCTAGGCATCGCACATGATTGTTACCTGTTGCAGAAAGAA 60

QY 3180 taaaaatgacataagaacaaactctcttcaaaaagtggagatttccaaatgtccccatatgc 3239
Db 61 TAAAAATGACATAAGCAAACTCATCTTCAAAAGTGGAGATTTCCAAATGTCCCGGTATGC 120

QY 3240 tgaatacc 3247
Db 121 TGAATACC 128

RESULT 9
US-08-925-743-7
; Sequence 7, Application US/08925743
; Patent No. 6054558
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,743
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/485,573
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; US-08-925-743-7
```

```
Query Match          1.8%; Score 126.4; DB 3; Length 128;
Best Local Similarity 99.2%; Pred. No. 1.2e-25;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3120 gggaggtgggctctctatccttaggcacgcactgattgttacctgttgcagaaagaa 3179
Db 1 GGGAGGTGGGCTCTCTGCTCATCTAGGCATCGCACATGATTGTTACCTGTTGCAGAAAGAA 60

QY 3180 taaaaatgacataagaacaaactcttcaaaaagtggagatttccaaatgtccccatatgc 3239
Db 61 TAAAAATGACATAAGCAAACTCATCTTCAAAAGTGGAGATTTCCAAATGTCCCGGTATGC 120

QY 3240 tgaatacc 3247
Db 121 TGAATACC 128

RESULT 10
US-08-925-767-7
; Sequence 7, Application US/08925767
; Patent No. 6225084
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,767
; FILING DATE: 09-SEPT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-097
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; US-08-925-767-7

Query Match          1.8%; Score 126.4; DB 4; Length 128;
Best Local Similarity 99.2%; Pred. No. 1.2e-25;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3120 gggaggtgggctctctatccttaggcacgcactgattgttacctgttgcagaaagaa 3179
|||||
```


Search completed: July 26, 2002, 01:37:45
Job time: 20054 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 19:34:16 ; Search time 12075.4 Seconds
(without alignments)
12047.710 Million cell updates/sec

Title: US-09-840-746-2

Perfect score: 6952

Sequence: 1 gttgatgaagaattgccg.....tactgggaagctatagggtg 6952

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_ov.*

5: gb_pat.*

6: gb_ph.*

7: gb_pl.*

8: gb_pr.*

9: gb_ro.*

10: gb_sts.*

11: gb_sy.*

12: gb_un.*

13: gb_vi.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_un.*

28: em_vl.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description

1	5140.4	73.9	5142	9	AB033063	AB033063 Homo sapi
2	5098.4	73.3	6407	6	AR065869	AR065869 Sequence
3	3508.6	50.5	189933	2	AC026342	AC026342 Homo sapi
4	3500.6	50.4	110000	2	AC092983_0	AC092983 Homo sapi
5	1438.8	20.7	1462	9	BC004539	BC004539 Homo sapi
6	1384.8	19.9	1397	9	BC014053	BC014053 Homo sapi
7	1359.8	19.6	110000	2	AC092983_0	AC092983 Homo sapi
8	663	9.5	118407	2	AC026629	AC026629 Homo sapi
9	610.8	8.8	118407	2	AC026629	AC026629 Homo sapi
10	387.4	5.6	400	6	AX245407	AX245407 Sequence
11	326.2	4.7	50921	2	AC092983_3	AC092983 Homo sapi
12	266.4	3.8	115766	2	AC069571	AC069571 Homo sapi
13	263.4	3.8	34308	9	AF389852	AF389852 Homo sapi
14	126.4	1.8	128	6	AR053542	AR053542 Sequence
15	126.4	1.8	128	6	AR080351	AR080351 Sequence
16	126.4	1.8	128	6	ARI48306	ARI48306 Sequence
17	112	1.6	112	11	G20489	G20489 human STS A
18	102.6	1.5	2008	9	AB062989	AB062989 Macaca fa
19	89.8	1.3	7218	6	I66494	I66494 Sequence 14
20	88	1.3	110000	2	AC092983_2	AC092983 Homo sapi
21	67.6	1.0	150110	2	AC097722	AC097722 Homo sapi
22	67.6	1.0	177475	2	AL626782	AL626782 Mus muscu
23	67	1.0	179262	2	AC079507	AC079507 Mus muscu
24	65.8	0.9	2384	3	PFAGAR	PFAGAR Plasmodium
25	65.6	0.9	397	11	G37798	G37798 GARP Plasm
26	65.2	0.9	56757	2	AC105856	AC105856 Rattus no
27	64.8	0.9	174591	2	AC103093	AC103093 Rattus no
28	64.4	0.9	197844	2	AC079176	AC079176 Homo sapi
29	63.6	0.9	156224	2	AC096059	AC096059 Rattus no
30	63.6	0.9	183272	2	AC094428	AC094428 Rattus no
31	63.4	0.9	91999	2	AC095962	AC095962 Rattus no
32	63.2	0.9	165077	10	AC084382	AC084382 Mus muscu
33	63	0.9	194472	2	AC096438	AC096438 Rattus no
34	62.8	0.9	166704	2	AC078821	AC078821 Homo sapi
35	62.6	0.9	184855	2	AL606744	AL606744 Mus muscu
36	62.6	0.9	209671	10	AC096777	AC096777 Mus muscu
37	62.4	0.9	126736	2	AC044788	AC044788 Homo sapi
38	62.4	0.9	182288	9	AC018362	AC018362 Homo sapi
39	62.4	0.9	203299	2	AP001178	AP001178 Homo sapi
40	62.4	0.9	244490	2	AC105495	AC105495 Rattus no
41	62.2	0.9	82973	2	AC105453	AC105453 Rattus no
42	62.2	0.9	170749	10	AL133159	AL133159 Mouse DNA
43	62	0.9	202904	2	AL603837	AL603837 Mus muscu
44	61.8	0.9	146438	9	HSBA191L9	AL117329 Human DNA
45	61.6	0.9	174990	10	AC027653	AC027653 Mus muscu

ALIGNMENTS

RESULT 1
AB033063
LOCUS
DEFINITION: Homo sapiens mRNA for KIAA1237 protein, partial cds.
ACCESSION: AB033063
VERSION: AB033063.1
KEYWORDS: 5142 bp mRNA linear PRI 11-NOV-1999
SOURCE: Homo sapiens brain cDNA to mRNA, clone_lib:pBluescriptII SK plus
clone:fh09696.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (sites)
Nagase,T., Ishikawa,K., Kikuno,R., Hirose,M., Nomura,N. and Ohara,O.

AUTHORS

Prediction of the coding sequences of unidentified human genes. XV.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro

DNA Res. 6 (5), 337-345 (1999)

JOURNAL

20039619

MEDLINE

2 (bases 1 to 5142)

REFERENCE

Ohara,O., Nagase,T. and Kikuno,R.
Direct Submission

||||| GCTTATCAGTGTGATGCGAGCGCGGGAGGTGGGCTCTGCTCATCTCTAGGCATCGC 1620
QY actgattgttacctgttgcagaaagataaaaaatgacataagcaaaactcatcttccaag 3212
Db ACTGATTTGTTACCTGTGTGCAGAAAGATAAAAAATGACATAAGCAAAATCATCTTCAAAAG 1680
QY tggagatttccaaatgtcccatatgctgaatacccccaaaaactcctgctcacaagaatg 3272
Db TGGAGATTTCCAAATGTCCCGGTATGCTGAATACCCCAAAATCCTCGCTCACAAAGATG 1740
QY gggccgagaagctattgaaatgcatgagaatggaagtaccaaaacacacctcctccagatgac 3332
Db GGGCCGAGAAAGCTATTCAAATGCATGAGAAATGGAAGTACCAAAACCTCTCTCCAGATGAC 1800
QY ggatgtactactcgctcacaagtgtaagaatccagaacttgaaacgaacgacctcta 3392
Db GGATGTGTACTTCGCTTCAAGTGTAAAGTGTAAAGAAATCCAGAACTTGAACGAACGACACTA 1860
QY ccgggctacactggactgcagatcacgcatctcttgcatcttcccgagacagtataa 3452
Db CCGGCTTTCATCAGTCATGAAAGCAGACAGAAAGAGACTACTTTAAGTCCAGGAGAGAGA 1980
QY ggactcatgtctgagccagtcacactggacctctgctcagagagccgacacagagag 3572
Db GGGACTCATTTGCTGTGAGCCAGTCACTCGGACCTCTGCTCAGAGGACCCACACAGAGG 2040
QY ctgcccagagatttgcggagccagctgagtggaagcaagaggaagagagagagagcatg 3632
Db CTGCGCCAGGATTTGCGGGAGCCACGCTGAGTGGCAAGCAGGAGGAGGAGGAGG 2100
QY cggggcgtagccagctggaggagacaggtggatgtggaaccacagcagctctcattcagc 3692
Db CCGGGCTGACACAGCTGGAGGAGACAGGTGGATGTGGAACACACAGGCTGCTCATTCAGC 2160
QY acctgtgttactgtgaacgtgaatgtggccagatatacaagagagctctctcagatga 3752
Db ACCCTTTGTTACTGTGAACGTGAATGTGGGCCAGTATCAAGAGAGTCTCTCTGAGTGA 2220
QY ctgacacatggcactggccacagggagactattagccaggcgacagaccactagacttcagt 3812
Db CTCACCATGGCATGCGACACAGGCGACTATTAGCCAGGCGACACCACTAGACTTCAGT 2280
QY gcaggagacctggtttccctctgcttgcactttagtaaaattgggtggaggtttcccttt 3872
Db GCAGGGACCTGGTTCCTTCCTTCTGCTTTCAGTAAATTTGGTGGGAGGTTTTCCTTTT 2340
QY ggaatctgttgagactgttccagaagaaggcttcccttcccgagacacttccatagcc 3932
Db GGATCTGTTTGAGACTGTCTCCAGAAAGAGGCTTCCTTTCCCGAGACACTTCATAGGC 2400
QY agcaattgtgtattcttcgagcaaaaactggctgttgtaattatttccctgcccagc 3992
Db AGCAATTTGCTGATTCATTTCGACAAATACTGGCTGTGTTAATTATTTCTCGCCAGC 2460
QY gctcgctgtaaacacaacagatgagatgagcgctaccactgaagctggaagatgtgccca 4052
Db GCCTCGCTGCTAAACACAGATGAGGATGAGCGTACCACCTGAAGCTGGAAGATGTCGCCA 2520
QY tgaacggacagtgctttcctatagtttctaggttgctttatgctacagtttccaagccag 4112
Db TTGAACGGACAGTGTTTTCATATATGTTTCTAGGTTGTCTTATGCTTACAGTTTCCAAGCCAG 2580
QY cccccagtgagaaatgttcaggcacccagacacaaactgcaatgtgttttttaagtca 4172
Db CCCCCAGGTGAGAAATGTGTAGGCGACCGCACAACTGCAATGTGTTTTTAAGTCA 2640
QY aggtgacacatgtatttaagatttttttaaaatctcttgcagttaaaatctcactttt 4232
|||||

Db 2641 AGGTGACACATGTATTTAAGATTTTTTTTTTAAAAATCTCTTTGACAGTTAAATCTCACTTTT 2700
QY tcaacaagcctgcatcagggcacaacaacttataatttggtttttagctggaggctcagca 4232
Db TCAAAACAGCCTGGATCAGGCGAAAACAACTATATTTGGTTTTAGCTGGAGGCTCAGCA 2760
QY ggcagattgcagcaggggggacacttttcatcscatgagggcccaagcctgggacctgggac 4352
Db GGCAGATTGAGGCGAGGGGGGACATTTTCATCCTATGAGGGCCAGCCTGGGGCCTGGGAC 2820
QY tctgataccactattgagggcagagcgatgcgtatgagggagaaatgccaactgaa 4412
Db TCTGATCACCATTTGTGGAGCCAGAGCGAGCTGCTATGGAGGAGAAATGTCAAACGTAA 2880
QY cgaagtttccaccacttagaagcagctgtttagacccctgacatggaatggttag 4472
Db CGCAGGTTTCACTCTAGAAAAGCAGCTGTTTGAAGCCCTGCAGCTGGATGTGGTTAG 2940
QY aggatggctgaaatagcgaggttagatttccctgcatacaacagtgctttggaaagctgtg 4532
Db AGGATGGCTGAATAGCGCAGGTTAGATTTCTCTGCATCAACAGTGCCTTTGGAAAGCTGTG 3000
QY tggattcctgaggaagacacagggcgagatggagccacacatgagtttgcctccagcgc 4592
Db TGGATTCTTGAGGAAGAACAGGGAGCCGAGATGGAGCCACACATGAGTTTGTCTCACCGGC 3060
QY tactgcagcacttgttaccagaaatctcatgtccaaaaacccatgtaaaactttcaacca 4652
Db TACTGCGAGCATTTGTATCCAGAAATCTCATGTCACAAAACCCCATGTAAACTTTCAACCA 3120
QY ctcaagctgttattcgcgtgaaagaaataacttttttctcaccagctcattgttacc 4712
Db CTCAAGCTGTTATTCGGCTGAAGAAATAACTTTTTTTTCTCACCCAGTCATTTGTACC 3180
QY tcttcattggctgtgtcgcacccctccagaaacgtggtttatacttccagtcagtggtgga 4772
Db TCTTCATATGCTGTGCGCACCCCTCCAGAAACGTGGTTATATTCTCAGTCAGTGTGGGA 3240
QY gaactgaagacttccggttggctgaggaactgaggttgacacttcgggaaggaagtcca 4832
Db GAATGAAGACTTCGGGTGTGTCGAGGAACGTAGGGTTGACCTTCGGGAAGGAAGTTCCA 3300
QY ctcatctatttattatgctgtgctgaggtcctgcagggagagacataccagtagtccgt 4892
Db CTCATCTTATTATTATGCGCTGTGATGTGGTCTCGCAGGGAGACATCCAGTACTCGGT 3360
QY gtctttaatgtccacctggggaactgtgttatttggccttcttggggatccctgggtttt 4952
Db GTCTTTAATTGCCACCTGGGGAACTGTGTTTATTTGGCCTTCTTTGGGCGCATCTGTGTTT 3420
QY ggaatgaagtagggggaatacagaggttaaaagaaattgtctccacctgaagcgggaagtc 5012
Db GGATGAAGTGAAGGGAATACAGAGTAAAGAAATGTGCTCCACCCCTGAAGCGGGAGTCC 3480
QY cgcttcacattctcggaaatgggtcagccactggggagcagttctgcccggcgatgggtg 5072
Db CGCTTCACATTTCTGGAATGGTGCAGCCACTGGGGAGACTTCTGCCCGGGCATGGTTG 3540
QY tttctcaagtcctctaaataataatccctattcttacaataatccttggccctgatggtt 5132
Db TTTCTTCAAGGTCCTCTAAATATAATCCCTATTCTTACATAATCCTTTGGCCCTGATGGTT 3600
QY ttaagcaagaactcctgtgcccatgggtctccaccactcaccatcaacctgctgtagcaa 5192
Db TTAAGCAAGAACTCCTGTGTCCTCATGGTCTCCACACTCCACATCACCCCTGCTAGCAA 3660
QY gactcctagtcagggaggtgcatttttagtttagttaatgacttaacttaccatgaagaaat 5252
Db GAGTCTCTAGTCAGGGAGGTGCATTTTAGTAGTTAAATTTGCACATTATCCATGAGATAAT 3720
QY aaaagaggaactgtttttatcatcagtggaaggttaacctaaaaatttcaaatgtgcgctttt 5312
Db AAAAGGAGAACTGTTTTATCAGTGGAGGCTAACCTTAAATTTCAAAAGTGTGCGCCTTTT 3780

```
Qy 5313 gaaatcttggccctctctctctgtagaaccaatgccccttcttggtccagggcctcgac 5372
|||||
Db 3781 GAAATCTTGGCCCTCTCTCTGTGTAGAACCAATGGCCCTTTGTGGCTCAGGCCCTCCGAC 3840

Qy 5373 ctaactggagagttctgagctctctgagctcactgagccacagactaggtctctggc 5432
|||||
Db 3841 CTAACCTGGAGAGTTCTGAGCTCTGTCAGCTCACCTGAGCCACAGACTAGGCTTCTTGGC 3900

Qy 5433 tccttcagagcatgctgctcaccaccagaaacccgagctgtggaagagccatgtagg 5492
|||||
Db 3901 TCCCTCCGACATGCTGTCTACCCCCAGAACCCGAGCTGTGGGAAGAGCCATGTAGG 3960

Qy 5493 gagctattccagagcatcacttccactccctcagctgacgtcacagctgacaaatca 5552
|||||
Db 3961 GAGCTATTCCAGGCATACACTTCCACTGCCCTTACGCTCAGCTCAGCTGACGAATCA 4020

Qy 5553 tctctctatcgagccagaaagacttcagctccacaaaaaagaagtgtctctgctcgaaa 5612
|||||
Db 4021 TCTCCTCTATCGGAGCCAGAAGACTTCAGCTCCACAAAATGAAGTGTCTGTCTCGTGA 4080

Qy 5613 cattcttgggaagaatcccaacatcgagaaaaaggtgtctctgtgagttccaaacaatgctt 5672
|||||
Db 4081 CATTTCTTGGGAAGAATCCCAACATCGAGAAACGGTGTCTGTGAGTTCCAACAATGCTT 4140

Qy 5673 ctgttctcatgggttctctcgatggagtgagattaaagtggtttattttgttgtctaa 5732
|||||
Db 4141 CTGTCTCATGGGTTCCTCCGTATGGAGTGGATTAAAGCTGTTTATTATTGTGTCTTAA 4200

Qy 5733 ctgagaaaaaaggaggcaccacaaaggttgaggtcacaagctctccacagtttccagg 5792
|||||
Db 4201 CTGAGAAAAAAGGAGGACCCACCAAGTTTGAGGTCAACAGTCTCCACAGTTTCCAGGA 4260

Qy 5793 ggcgttgggggtgggaagaccctccagagcatgaggtctaaaggggacatgagtaaa 5852
|||||
Db 4261 GGCCTTGGGGGTGGGAAGGCACCTCCAGAGCATAGAGTCTTAAGGGGACATGAGTAAA 4320

Qy 5853 gcatgtctgtgaccagtgaggaaggagagggccagctgcactctcagcggggttcccta 5912
|||||
Db 4321 GCATGTCTGTGACCCAGTGAAGGAGGAGAGCCAGCTGCACCTCTCCAGGGGTTCCTA 4380

Qy 5913 gctgcagaagggtccgccttaggcggagggaacacctgatagcagaaggcctggat 5972
|||||
Db 4381 GCTGCAGAAGGTCCTCCCTTAGGCCGAGGGGAACACCTGATAGCAGAAGAGCCTGGAT 4440

Qy 5973 qcacactgcagccagagctctccgcccagacacagatgctcattgacccctcac 6032
|||||
Db 4441 GCACCTTGGCACCCGAGGCTCTCCGCCAGACACAGTGTCCATGTACGCCCTCGAC 4500

Qy 6033 ctgggggtgtgtgattcagtgacagatgccaacactctgcacaa tatcccacagatgg 6092
|||||
Db 4501 CTGGGTGTGTGATTACGTGCAGATGCCAAATCTCTCACCAATATCCACAGATGG 4560

Qy 6093 ggaaggtgcagagaaagggaagtgtgtgtaactgtcacaagatgcttaaacctcca 6152
|||||
Db 4561 GGAAGGTGCAGAGAGGGGCAAGTGTATGTAACTGCTCAAGAGATGCTTAACCTCCA 4620

Qy 6153 tagagagagccgggcaggggcatctgtgtcccgtcacacactgcagcaggggaag 6212
|||||
Db 4621 TAGAGAGAGCCGGGCGCAGGGGCATCTGTGTCTCCCTCACACATTCACAGAGGGAAG 4680

Qy 6213 gtggctggctggctcccttgcatcaagtgtgtgtgttttaagctccagagggtcttattggc 6272
|||||
Db 4681 GTGGCTGGCTGGCTCCCTGGCATCAGTGTGTGTGTTTAAAGTCTCAGAGGGTCTTATTGCC 4740

Qy 6273 atgtctttctctgccccttagcagcagcactaaggccctggagctgttcttcttaaggc 6332
|||||
Db 4741 ATTGTCTTTTCCCTGTGCCCTTAGCCAGGCTTAAGGCCCTTGGAGTCTGTCTTTTAGCGC 4800

Qy 6333 gatgaactgcagtctctaccatgaccagggctctctggcaaggctctctcacagtatcctt 6392
|||||
Db 4801 GATGAAGTGCATGCTCTCTACCATGACACGGCTCTGGGCAAGGCTCTCTACAGTATCTT 4860
```

```
Qy 6393 gaggtggcgatggaagtcccatcttcaggttacagaaaccttcagagaggataaata 6452
|||||
Db 4861 GAGAGGTGGCATGGAAGTGCCCATTTCTCAGGTACAGAAACCTTCAGAGAGGATAAATA 4920

Qy 6453 gcttgccctgtagaagcaggactgaaaccccttgcgccttgactccccccagctactctgc 6512
|||||
Db 4921 GCTTGCCTCTAGAAAGCAGGACTGAAACCCCTTGTCCGCTGACTCCCCAGCTACTCTGC 4980

Qy 6513 ccactgtagccctgccttactgtcctgcacacccctcaccatcctgtataccctaaa 6572
|||||
Db 4981 CCACGTAGCCCCCTGCTTACTGTCTGGCACACCCTTACCATCTGTATACCTTAAA 5040

Qy 6573 tatcaagagggcaagagagaaagggtttaagaataagttattttttaagggaacctta 6632
|||||
Db 5041 TATCAAGAGGGCAAGAGAGAAAGGGCTTTAAAGATAAGTTATTTTAAAGGAACCTTA 5100

Qy 6633 atattattttaagaagtaacaaaatttagtgacgtgaatgc 6674
|||||
Db 5101 ATATTATTTTAAAGAGTAACCAAAATTAGTCACGTGAAATGC 5142

RESULT 2
AR065869
LOCUS
DEFINITION Sequence 7 from patent US 5849578.
ACCESSION AR065869
VERSION AR065869.1 GI:5996085
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6407)
AUTHORS Falb,D.A.
TITLE Compositions and methods for the treatment and diagnosis of cardiovascular using RHD528 as a target
JOURNAL Patent: US 5849578-A 7 15-DEC-1998;
FEATURES
    Location/Qualifiers
        source
            1..6407
            /organism="unknown"
BASE COUNT 1594 a 1762 c 1520 g 1509 t 22 others
ORIGIN

Query Match 73.3%; Score 5098.4; DB 6; Length 6407;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 5412; Conservative 21; Mismatches 8; Indels 301; Gaps 2;

Qy 1 gtcgatgaagaattccgcgttttccaacaaagagtggaaacgctcgagatggaac 60
|||||
Db 666 GTTCGATGAAGAATTCCGCTTTTCAACAAAGAGTGAACAGCCTCGGAGATGGGAAC 725

Qy 61 agagagggcatgggctgcagaagaatgactgtcacagccacagagccacccttc 120
|||||
Db 726 AGAGAGGGCATGGGGCTGTCAAGAAGTGGACTGTGCACAGCCAGAGGCCACACTTC 785

Qy 121 ggtttgagccctcccttctctctgttggagatgggagagctgaccagcctcttag 180
|||||
Db 786 GGCTTGAGCCCGTCCCTTTCTCTCTTTGGAGATGGGAGCTGACACGCTTCTTAG 845

Qy 181 gaagagaaattcctcaggaaccagatctctctggctgctcattctcacaggacagcttc 240
|||||
Db 846 GAAGAGAAATTCCTCAGSACCAGATCTCTCTGGCTGCATTTCTACAGGACAGCAGTTC 905

Qy 241 ctctctctcttagaccttctcaccctctgaaagtacagagaagcttaacaaactcac 300
|||||
Db 906 CTCTCTCTCTTAGACCTTTCCCTCACCTTCTGAAAGTACAGAGAAGCTTAACAACCTCAC 965

Qy 301 tggcctccagagctcctcagtcagtcacaaacaaacaaatcatgttctaccgtgttccac 360
|||||
Db 966 TGGCCTCCAGAGCTCCTCAGTCAGTCAGTCAAAACAAACAAATGCATGTTGCTACCGTGTTCAC 1025

Qy 361 tgatgtggccccgagaacgctggagatcttgcaggtcagctctggagacctgtgagcaagac 420
|||||
```

Db 1026 TGATGGTGGCCGAGAACGCTGCGATCTTTACGGTCAGTCTGGGACCTGTGACGAAGAC 1085
QY 421 agaaggctccccaaagactccagaattgccaagactctcatctcctcagctcctcttctcacc 480
Db 1086 AGAAGGCTTCCCCAAGACTCCAGAATTGGCAGGACTTCATCTCAGCTCCTTTTTCACC 1145
QY 481 ctctcgagtggaatcgagaagaacagtagagtaactgggaatccacaggggatgaggaatt 540
Db 1146 CTCTGCACTGGATCGAAGAGAACAGTAGAGTAACCTGGGAATCCAGGGGATGAGGAATT 1205
QY 541 cattgaaccatccacagaaataaattggacttaogtcttttgcg-tggcaaaatgattc 599
Db 1206 CAATTGAACCATCCACAGAAATGAATTTGGACTTACGCTTTGGCTTGGCAAAATGATTC 1265
QY 600 cccaaccttggagaacatcagcttgcagcagctctcagagtgcaaaatggaagctcccat 659
Db 1266 CCCAACCTTTGGAGAACATCAGCTTGCAGCAGCTCTGAGGTGCAAAATGGAATGCCAT 1325
QY 660 gtctcagactgagactgtctaggtcagtcgcacccatgagagtgagagatcactgc 719
Db 1326 GTCTCAGACTGAGACTGTGTCTAGGTGAGTCCGACCCATGAGAGGTGAGAGATCACTGC 1385
QY 720 aactggctcttgaccacagcaacacatctgcagatgtgacagggagctctgtctcata 779
Db 1386 ACACGTGGCTCTTGACCAACACAGCACACATCTGCAGATGTGCAGGAAGCTCTGCTTCATA 1445
QY 780 tectgaagtgtaagtcttcagtggttgaccaggtctcagactctactgtacagctgg 839
Db 1446 TCCTGAAGGTGTGAATGCTTTCAGTGTGACCCAGTTCACAGACTCTACTGTACAGTCTGG 1505
QY 840 aggaagtcacacagcattggagatagaggttattcagagctcttcacatctctc 899
Db 1506 AGGAAGTCACACAGCATTGGAGATAGGAGTTATTACAGACTTTCATCTACATCTTCTC 1565
QY 900 ggaagcttgaattcatcagcaccacgaogtggagaacttcaa----- 940
Db 1566 GGAAGCTTGAATTCATCAGCACCACGCTGGAGAACGTTCAACCTTGGAGACAGCCGAGA 1625
QY 941 ----- 940
Db 1626 GCCAGCCCAAGCACTAGGTGACAGTTCGCCCAATGACAGAGGACAGGACTTCTGGGGTGCC 1685
QY 941 ----- 940
Db 1686 CTCTCTCGGCACCCACACCTTGGCTACTGTCACTGTGGAACGGGGAGCGACACTGCGGTC 1745
QY 941 ----- 940
Db 1746 TGTACCCCTCACCAACACCAGCATGAGCAGGACTTCTTGGGGAAGCAGGCGCTGCAGC 1805
QY 941 ----- 940
Db 1806 GGCCATGCCCCAAGAAACAGAGGGTGCTCTCTGACAGTAAACGTGACGGACGACATGGG 1865
QY 941 -----tcgtgtggattagctacgg 959
Db 1866 CTGGGTCTCAGGTCAGTGCCGCCCTCCACTGCATCGACTCGGAGTCTGGGGATTAGCTACGG 1925
QY 960 tcaagtgcgtggaacagctattgaacaaaggacttccagcgaccacacagaccaccta 1019
Db 1926 TCAAGTGCCTGGCACAGCTATTGAACAAAGGACTTCCAGGAGCCACACAGACCACCTA 1985
QY 1020 cctgtcatctactttcaccaaaaggagaacggcggttactgtcattatcacagataacagttc 1079
Db 1986 CTTGTATCTACTTTCACCAAGAGGAGAACGGGGTTACTGTCCATTACAGATAACAGTTC 2045
QY 1080 atctcagacattgtgagagctcaactctctatatataaactcctcaactcttcacattc 1139
Db 2046 ATCCTCAGACATTTGGAGAGCTCAACTTCTTATATTAATAATCTCAAACTCTTCACATTC 2105
QY 1140 agagtattctctcttctcagactgagagaggaagtaacatctcatctcctatgacgg 1199
Db 2106 AGAGTATCTCTCTTTTCTCATGCTCAGACTGAGAGAGTAAACATCTCATCTATGACGG 2165

QY 1200 ggaatatgctcagcctctactgagtcgacagttctctgcatacatccaaaccttcgctcta 1259
Db 2166 GGAATATGCTCAGCCTTTCTACTGAGTCGCCAGTTCTGCATACATCCAACCTTCGTCCTA 2225
QY 1260 caacccaccatataatgctcgaaacacttgggttctctggacactgatgctgagttgt 1319
Db 2226 CACACCCACCATTAAATATGCGAACACTTCCGGTGTCTTGACACTGATGCTGAGTTGT 2285
QY 1320 tagtact 1379
Db 2286 TAGTGACT 2345
QY 1380 gcctctccctctgtgcacaatcccaactttattttcaatattttacatacaacag 1439
Db 2346 GCCTCTGCCCTCTGTGTGCAAAATCCACCAATTTATTTTCATCAATTTTACCATCAACAG 2405
QY 1440 ggctctgtgcattactactaaagtctactctgtatgcattccacacatggtctctctcacc 1499
Db 2406 GGCTCTGTGCATCTACTAAAGTCTACCTCTGATGCATCCACACATGGTCTCTCTCACC 2465
QY 1500 atcaacttaccagttactcttaacagacatctacatctgccccacttctgtctcacaaac 1559
Db 2466 ATCACCTTTACCAAGTATCTTAACGACATCTACATCTGCCCACTTTCTGTCTCACAAC 2525
QY 1560 aacttgcacagtcattcttaccctgtctgccccaggggcaaggagactcctgtgac 1619
Db 2526 AACCTTGCCACAGTCATCTTCTACCCCTGTCTGCCAGGGCAAGGGAGACTCTCTGTGAC 2585
QY 1620 ttcatcttcagacatcaacaatgacatcttcatgacaatgctccatagtagtcaaaactgc 1679
Db 2586 TTCAATTTACAGACATCAACAATGACATCATTCATGACAATGCTCCATAGTAGTCAAACTGC 2645
QY 1680 agaccttaagcagcagagcaccacacacacacacacacacacacacacacacacacac 1739
Db 2646 AGACCTTAAGAGCCAGAGCAGCACCACACACCAAGAGAAAGTCAATTACAGAACTCAAACTCACC 2705
QY 1740 aagctgtgtctctgcccacagagtcacacaaagctgtaacacaaactctcttggc 1799
Db 2706 AAGCTGTGTCTGTGCCACAGAGTCCACAAAGCTGTAAACAACAACTCTCTCTTTGCC 2765
QY 1800 tccatctcttaacagagtcctccacagacaaaccttccagccacacacacacacacacacttagc 1859
Db 2766 TCCATCTTAAACAGAGTCTCCACAGAGCAAACTTCCAGGCCACAAGCAACTTAGC 2825
QY 1860 acaaatgtctccaaacttccaaactaacttctgaagacactctcagctctcttctgaacac 1919
Db 2826 ACAAAATGCTCCAACTTTCACTACACTAGCATTTCTGAAGACCTCTCAGGCTCTCTTATGACCAC 2885
QY 1920 tctggcaccctgtcgaagcagcagctctgttcaactggccctatagcgtacagactac 1979
Db 2886 TCCTGGCACCCCTGTCAAGCAGCAGCATCTCTGTCTCCTTGGCCCTATAGCCGTACAGACTAC 2945
QY 1980 agctggaaaaacagctctcgtgacctcctgaaataactagttctcctcaaatctcacaaga 2039
Db 2946 AGCTGGAACACAGCTCTCGCTGACCCATCTCGAATACTAGTTCTCTCAATCTCAACAGA 3005
QY 2040 aggtggcatcagcacagaaaggaaacccagtgattgttgatgtctaccactggattgatccc 2099
Db 3006 AGGTGGCATCAGCACAGAAAGGAACCGAGTGATTGTGGATGCTACCACTGGATTGATCCC 3065
QY 2100 tttagaccagtgtaaccacatcagcaaaaagaaatgacacaaaagcttggcgttacacagaga 2159
Db 3066 TTTGACCAGTGTACCCACATCAGCAAAAGAAATGACCACAAAGCTTGGCGTTACAGCAGA 3125
QY 2160 gtacagcccagctcagcttccctcggaacatctcctctctcccaaaacacagttgttc 2219
Db 3126 GTACAGCCAGCTTCAGTTCCCTCGGAACATCTCTCTCTCTCCCAAAACACAGTTGTTTC 3185
QY 2220 cagggctgaagacttggtctcccaaatctgcacacttctgtgttcagagcagcacacagtc 2279
Db 3186 CACGGCTGAAGACTTGGCTCCCAAAATCTGCCACCTTTGCTTTCAGAGCAGCAGCACAGTC 3245

Db	5406	GCTTGTGACCCCTGACACCTGGATGGTTAGAGGATGGCTGAATAGSCAGGTTAGA	5465
QY	4500	tttcctgcatcaacagctgtttggaagctgtgtgattcctgaggaagaacagggagcc	4559
Db	5466	TTTCTCTGATCAACAGTGTCTTGGGAASCTGTGTGATTCTCTGAGGAAGAACAGGGAGCC	5525
QY	4560	gagatgagccacacatgagttgtgtcaccgggtactgcagcactttgtaccacagaatct	4619
Db	5526	GAGATGGAGCCACACATGAATTGCTCACCGCTACTGCAGCACTTGTACCCAGAATCT	5585
QY	4620	catgtccacaaacccatgtaaacctttcaaccactcaaaagctgtttatttgcgtgaagaa	4679
Db	5586	CATGTCCACAAACCCCATGTGTAACCTTTCAACCACTCAAGAGTGTATTTCGGTGAAGAA	5645
QY	4680	ataacttttttctcaccagcatattgtacctcttcatatgctgtgtgcacacctcc	4739
Db	5646	ATAACTTTTATTTCTACCCAGTCAATTGTGACCTCTTCATATGATGTGCGCACCTCC	5705
QY	4740	agaaacgtggttacttccagtcagtggtggagaactgaagacttccggttggctgag	4799
Db	5706	AGAAAGTGTGTTACTKCACTGAGTGGGAGAACTGAAGACTTCCGGTTGGTCAGG	5765
QY	4800	aactgaggggtgaccttcgggaaggatgccactcactcatttatttgcctgtgat	4859
Db	5766	AACGTGAGGTTGACTTCGGGAAGAAATCCCACTCATCTATTATTATCCCTGTGATG	5825
QY	4860	tgggtctccagggagacatcagctactggtgttcttaattgcacctggggaactgt	4919
Db	5826	TGGGTCTGCGCAGGGAGACATCCAGTACTCGGTGTCTKTAATTCGCCACCTGGGAACTGT	5885
QY	4920	gtttattggcctcttttggggcctcctggttttggatgaagtgaagggaatacacagagta	4979
Db	5886	GTTTATTGGCTCTTTTGGGGCATCTCGKTTGGATGAAGTGAAGGGAATACAGAGTA	5945
QY	4980	aagaattgtctccacctgaagcggggagtcctcgttccacttcttggaatggtgcag	5039
Db	5946	AAAGAATGTCTCACCTCAAGCGGGAGTCCCGGTTTCACATTTCTGGAATGTTGCAG	6005
QY	5040	ccactgggacagttctgccccgggcatggtgtttcttctcaaggctcctcaataataac	5099
Db	6006	CCACTGGGGACAGTTCGCCCCGGGATGTTGTTCTTCAAGGTCTCTAAATATAATC	6065
QY	5100	cctattctacataactcttggccctgatggttttaagcaagaactcctctgtccccatgg	5159
Db	6066	CCATTCTTACATAATCTTGGCCCTGATGTTTAAAGCAAGAACTCTCTGTCOMATGG	6125
QY	5160	tctccaccactcaccatcaccctgctgtgtagcaagagtcctagtcaggggaggtgcattt	5219
Db	6126	TCCTCCACCACTCACCATCACCTGCTGTAGCAAGAGTCTTAGTCAGGGAGGTGCATTT	6185
QY	5220	agtagttaattgacattatccatgagataataaaaggagaactgtttttatcagtga	5279
Db	6186	AGTAGTTACATTGCATTAATCCATGATGATGTTTAAAGCAAGAACTCTCTGTCAGTGA	6245
QY	5280	ggctaaacaaatttcaaaagtgcctcttttgaactcttggcctctctctctctgtaga	5339
Db	6246	GGCTAACCTAAATTTCAAGTGTGCTTTTGAATCTTGGGCTCTCTCTCTGTAGA	6305
QY	5340	accaatggcccttgttgctcagcggcctgcacctaactggagagttctgagctctgca	5399
Db	6306	ACCAATGCCCTTTGTGGTTCAGCGGCTCGCACCTAACTGGAGAGTCTGAGCTCTGCA	6365
QY	5400	gtcactgagccacacagactaggtcttcttggtctctctccgc	5441
Db	6366	GCTCACCTGAGCCACAGACTAGGCTTCTTGGCTCTCTCTCCGC	6407

RESULT 3
AC026342/c 189933 bp DNA linear HTG 31-JAN-2002
LOCUS AC026342 Homo sapiens chromosome 3 clone RP11-559L12, WORKING DRAFT
DEFINITION SEQUENCE, 7 unordered pieces.
ACCESSION AC026342

VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

AC026342.21 GI:18449521
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 189933)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhaq,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,D.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokoko,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 189933)
Worley,K.C.
Direct Submission
Submitted (22-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 31, 2002 this sequence version replaced gi:15625920.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HARC
Center clone name: RP11-559L12
----- Summary Statistics
Sequencing vector: M13: L08821
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329First call to
findPhrapList

Consensus quality: 189260 bases at least Q40
Consensus quality: 191512 bases at least Q30
Consensus quality: 192994 bases at least Q20

Estimated insert size: 189583; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 6.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 82312: contig of 82312 bp in length
 * 82313 82412: gap of unknown length
 * 82413 117485: contig of 35073 bp in length
 * 117486 117585: gap of unknown length
 * 117586 147623: contig of 30038 bp in length
 * 147624 147723: gap of unknown length
 * 147724 166827: contig of 19104 bp in length
 * 166828 166927: gap of unknown length
 * 166928 179740: contig of 12813 bp in length
 * 179741 179840: gap of unknown length
 * 179841 186370: contig of 6530 bp in length
 * 186371 186470: gap of unknown length
 * 186471 189933: contig of 3463 bp in length.

FEATURES

Location/Qualifiers
 source
 1. .189933
 /organism:"Homo sapiens"
 /db_xref:"taxon:9606"
 /chromosome:"3"
 /clone:"RP11-559L12"

BASE COUNT 54758 a 42732 c 41678 g 50156 t 609 others
 ORIGIN

Query Match 50.5%; Score 3508.6; DB 2; Length 189933;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 3553; Conservative 0; Mismatches 19; Indels 3; Gaps 3;
 QY 3342 ctactgcctacagtgtaaggatcaggaactcaggaactgaacgaacggagctctaccggccta 3401
 Db 155141 CAACCGCCCTACAAAGTGAAGGAATCAGAACTTGAAACGAACGGACTCTACCCGGCCTA 155082
 QY 3402 cactggagctccaggatcagcgattcttgcatttcccccggagcagataaacccgtcttt 3461
 Db 155081 CACTGGACTCCAGGATCAGCGCATCTTGCATTTCCCGGACAGTATAACCCGCTTT 155022
 QY 3462 catcagtga tgaagcagaagaagagactacttttaagtccagagagagagactcat 3521
 Db 155021 CATCAGTGTATGAAGCAGAAAGAGAGACTACTTTTAAGTCCAGGAGAGAGGGACTCAT 154962
 QY 3522 tgcctgagccagctacctggaccctctgctcagagaccgagaccgagaggtcgccca 3581
 Db 154961 TGTCTGAGCCAGTCACTGGGACCCTCTGCTCAGAGGACCGACACGAGGCTGCGGCCA 154902
 QY 3582 ggattgtcgggagccagctgagtggaagcagaggaagagggagcgatcgggcgctg 3641
 Db 154901 GGATTTGTCGGGAGCCAGCTGAGTGGCAAGCAGGAGAGGAGGATCGCGGCGTG 154842
 QY 3642 acccagtgagagagacagtgatgtggaaccacacaggtgctcattccagcactttgtt 3701
 Db 154841 ACCACAGTGAGGAGACAGGTGGATGTGGAAACACAGGCTGCTCAATCAGCACCTTTGTT 154782
 QY 3702 gttactgtgaactgaatgtggcgccagatcatcaagagagatctctctcagtgactgcaccat 3761
 Db 154781 GTTACTGTGAACGTGAATGTGGGCCAGTATCAAGAGAGTCTCTCTGAGTGACTGCACCAT 154722
 QY 3762 ggcactggaccagggcgactattagccagggcgagaccactagactcagtcagggacc 3821
 Db 154721 GGCACGTGGCACACGGGGACTATTAGCCAGGGCAGACCATAGACTTCAGTCAGTGGGAGACC 154662

QY 3822 tggtttccctctgcttgcacttttagtaaatgggtgggaggttctctcttttggatctgtt 3881
 Db 154661 TGGTTTCCCTTCGTTGCACCTTAGTAATGGTGGGAGGTTTCCTTTTGGATCTGTT 154602
 QY 3882 ttgagactgtccagaaaaaaggcttcccttcccgagacacacttccatagcagcaatttg 3941
 Db 154601 TTGAGACTGTTCCAGAAAGAAAGGCTTCTTTCCCGAGACACTTCCATATAGCAGCAATTG 154542
 QY 3942 gtattcatttcagcaaaaataactggctgttaattatttcttcccgagcgtcggtg 4001
 Db 154541 GTATTCTATTGTCAGCAAAATACTGGCTTGTTAATATTTCCTGCCACGCGCTGGGTG 154482
 QY 4002 ctaaacacagatgagatgagcgtaaccactgaagctgaagatgtcgccattgaaacgga 4061
 Db 154481 CTRAACAAACAGATGAGGATGAGGCTACCACTGAAGTCTGAAGATGTGCCCATTTCAACGGA 154422
 QY 4062 cagtgttttcatatgtttcttagttgtcttatgctacagtttccaaagccagcccccacag 4121
 Db 154421 CAGTGTTTTTCATATGTTTCTTAGTGTGCTTATGCTACAGTTTCCAAAGCCAGCCCCACAG 154362
 QY 4122 tgaggaaatgtgtgagccacgcacacaaactgcaaatgtgtttttaaagtcgaagtgcac 4181
 Db 154361 TGAGGAATGTGTGAGCCACCGCACACAACCTGCAATGTGTTTAAAGTCAAGGTGACAC 154302
 QY 4182 atgtatttaagattttttttaaatactctttgcagttaaatctcacttttttcaaacaa 4241
 Db 154301 ATGTATTTAAGATTTTTTTTTTAAATCTCTTGCAGTTAAATCTCACTTTTTCAAACAAG 154242
 QY 4242 cctggatcagggcacaaacacttatattgttttagctgagagctcagcagcgagattg 4301
 Db 154241 CCTGGATCAGGGCAAAACAACCTTATATTGTTTGTAGTCTGAGGCTCAGCAGGAGATG 154182
 QY 4302 caggcaggggggcacttttccatcatgagggcccagcctgggctgggactctgacac 4361
 Db 154181 CAGCAGGGGGGCACCTTTTCATCCATGAGGCCCGCAGCTGGGGCTGGGACTC-GATCAC 154123
 QY 4362 cattgtgagggccagggcagctgctgctatggagagagaaatgtcaaacctgaacgcagttt 4421
 Db 154122 CATTTGTGGAGCCAGAGGAGCAGCTGCGTATGGAGGAGAAATGTCAAACGTGAACGCGAGTTT 154063
 QY 4422 caccactctaggaaagcagcttattgagccctcagctgagatgtgttagagagatggg 4481
 Db 154062 CACCACCTCTAGGAAAGCAGCTTGTGTAGCCCTTCAGCTGGATGTGTTAGAGGGATGGG 154003
 QY 4482 ctgaataggcaggttagatttcttcgcatcaacagtgctttgggaagctgtgtggtatcct 4541
 Db 154002 CTGAATAGGCAGGTAGATTTCCTGCATCAACAGTGTCTTGGGAAGCTGTGTGATTCCT 153943
 QY 4542 gaggaaagacagggagccgagatggagccacacatgagtttgcaccgctactgcagc 4601
 Db 153942 GAGGAAGAACAGGGAGCGAGATGGAGCCACACATGATGTTTGTCTACCCGCTACTGCAGC 153883
 QY 4602 acttctgaccagaatctcatgtcccaaaaccccatgaaacttcaacacacaaagct 4661
 Db 153882 ACTTTGTACCCAGAACTCATGTGCCAAACCCCATGTAAACTTCAACCCACTCAAGACT 153823
 QY 4662 gttatttcggctgaagaaataaacttttttctcaccagctcatattgactctctcatat 4721
 Db 153822 GTTTATTTCGGCTGAAGAAATAAATTTTTCATCCACCGATCATTTGTACTCTTCATAT 153763
 QY 4722 ggctgtgtcagccctccagaaacgtgggttatacttccagtcagtgctgggagaaactgaag 4781
 Db 153762 GGTCTGTCCACCCCTCCAGAAACGTGGTTATATCTTCAGTCAAGTGTGGGAACTGAAG 153703
 QY 4782 acttccggttgcgaggaactgaggttgaccttggggaagaagttccacatcatcta 4841
 Db 153702 ACTTCCGGTTGGTCGAGGAACTGAGGTTGACCTTCGGGAAGGAAGTTCACATCATCTTA 153643
 QY 4842 ttattatgcctgtgatgtgggtcctgcagggagagacatccagctactcgtgtcttcaat 4901
 Db 153642 TTTATTATGCTGTGATGTGGGTCTGCCAGGAGACATCCAGTACTCGGTGCTTTAAT 153583
 QY 4902 tgcacctggggaactgtgtttatttgccotttcttggggcatacctcgtgttttggaagaagt 4961

Db	152502	GCACCCGAGGCTCTCGCCCCAGACACAGTGTCTCCATGTTCAGCCCCGACCTTGGGGTGT	152443
Qy	6042	gtgattcaagtgcacagatgcacaaaccttgacccaataccacagatgggggaaggtg	6101
Db	152442	GTGATTACGTGCACAGATGCCAACAATCCGTGCACCAATATCCACAGATCGGGGAGAGGTG	152383
Qy	6102	agagaaaggggcaagtgatgtgtaactgctcaagagatgcttaaacctccatagagaga	6161
Db	152382	AGAGGAAGGGCAAGTGATGTGTAACTGCTCAAGAGATGCTTAAACCTCCATAGAGAGGA	152323
Qy	6162	gccggggcgacgggcatctgtgtgcccgtaacacactgcagcagggaaaggtggcgtg	6221
Db	152322	GCCGGGCGCAGGGGCATCTGTGTCTCCCGTCACACACTGCAGCAGGAAGGGTGGTGCC	152263
Qy	6222	tggtccctgcatacaatggtgttggtttaagctccagaggtcttatgcattgcttt	6281
Db	152262	TGGCTCCCTGGCATAGTGGTTGGTTTAACTCCAGAGGCTCTATTGCCATTGTCTTT	152203
Qy	6282	tcctctgccccttgagccagcctaaggccctggagctgtttctttagcgcgatgaactg	6341
Db	152202	TCCCTGCCCTTCAGCCAGCCTAAGGCCCTGGAGTCTGTTCTTTAGGCGCGATGAAC TG	152143
Qy	6342	acatgctcctaccatgaccaggctctgtggcaaggtcctccacagtatccttgagaggtg	6401
Db	152142	ACATGCTCTTACCATGACCAAGCTCTGGGAAGGCTCCTCACAGTATCTCTTGAGAGGTGG	152083
Qy	6402	gcattgaagtgcattctcaggtacagaaaccttcagagaagataaatagcttgcct	6461
Db	152082	GCATGGAAGTGCCCATTTCTCAGGTACAGAAACCTTCAGAGAGGATAAATAGCTGCCCT	152023
Qy	6462	gtagaagcaggaactgaaaccttgcgcctgactccccagctactctgccactgtag	6521
Db	152022	GTAGAAGCAGGACTGAACCCCTTGCTCTGCCTGACTCTCCCCAGCTACTCTGCCACTGTAG	151963
Qy	6522	ccccctgcttactgtcctctggcacaccctccacactcctgtataccttaaatcaaga	6581
Db	151962	CCCCCTGCTTACCTGTCTGTCACACACCCCTCACCATCTGTATACCTTAAATATCAAGA	151903
Qy	6582	ggcaacagagaaaggctttaagataaagttattttttaaggaaccttaattattt	6641
Db	151902	GGGCAAGAGAGAAAGGCTTTAAAGATAAGTTATTTTTTAAAGAACCTTAAATATTATT	151843
Qy	6642	ttaagaagtaaccaaaattagtgacgtgaaatgc-aaaaaaaaaaaaaaaaaagtcgact	6700
Db	151842	TTAAGAAGTAAACCAATTAGTACGCTGAATGCANANANANANANANANANANANANANAN	151783
Qy	6701	acctttttgaaaaatgtctttccagatgtttttttatatgtaatctctagacacttgtcat	6760
Db	151782	ACCCCTTTTGAAAAATGCTTTTCAGATTGTTTTTTTATATGTAATCTTAGACACTTGTTCAT	151723
Qy	6761	taagaaataagtgtgcgtgtgctcagcaagaagcaactggcagctggcctttggtat	6820
Db	151722	TAAAGAAATAGTGCTTGGCTTGTGCTCAGGAAGGACCACTGGCAGGTGGCTTTGGTAT	151663
Qy	6821	aggaagtgaaggcaaggacctgggtttctgcagaagtgcgcgtcagacttaccctccatc	6880
Db	151662	AGGAAGTGAAGGCAAGGACCTGGGTTCTTGACAAAGTGCATGACATACACATCCATC	151603
Qy	6881	tggagagctgggtgttggctccctgggttagggc	6915
Db	151602	T-GAGAGCTGCTTGCTTGTGTTTCCGCTGGTAAGGCC	151569

RESULT	4	Sequence split into 4 fragments	LOCUS	AC092983	Accession	AC092983
AC092983_0						
WPCOMMENT						
		Fragment Name	Begin	End		
		AC092983_0	1	110000		
		AC092983_1	100001	210000		
		AC092983_2	200001	310000		
		AC092983_3	300001	350921		
LOCUS	AC092983			350921	bp	HTG
						linear
						DNA


```
* 289576 289675: gap of unknown length
* 289676 294061: contig of 4386 bp in length
* 294062 294161: gap of unknown length
* 294162 298031: contig of 3870 bp in length
* 298032 298131: gap of unknown length
* 298132 303266: contig of 5135 bp in length
* 303267 303367: gap of unknown length
* 303368 305907: contig of 2541 bp in length
* 305908 306007: gap of unknown length
* 306008 308221: contig of 2214 bp in length
* 308222 308321: gap of unknown length
* 308322 310563: contig of 2242 bp in length
* 310564 310663: gap of unknown length
* 310664 313552: contig of 2889 bp in length
* 313553 313652: gap of unknown length
* 313653 316728: contig of 3076 bp in length
* 316729 316828: gap of unknown length
* 316829 319666: contig of 2838 bp in length
* 319667 319766: gap of unknown length
* 319767 323675: contig of 3909 bp in length
* 323676 323775: gap of unknown length
* 323776 327559: contig of 3784 bp in length
* 327560 327659: gap of unknown length
* 327660 331218: contig of 3559 bp in length
* 331219 331318: gap of unknown length
* 331319 333580: contig of 2262 bp in length
* 333581 333680: gap of unknown length
* 333681 336280: contig of 2600 bp in length
* 336281 336380: gap of unknown length
* 336381 338607: contig of 2227 bp in length
* 338608 338707: gap of unknown length
* 338708 341633: contig of 2926 bp in length
* 341634 341733: gap of unknown length
* 341734 344669: contig of 2936 bp in length
* 344670 344769: gap of unknown length
* 344770 348238: contig of 3469 bp in length
* 348239 348338: gap of unknown length
* 348339 350921: contig of 2583 bp in length.
FEATURES             Location/Qualifiers
     source            1..350921
     organism="Homo sapiens"
     db_xref="taxon:9606"
     chromosome="3q"
     clone="RP11-717C3"
BASE COUNT      97816 a 75489 c 76879 g 95821 t 4916 others
ORIGIN
Query Match      50.4%; Score 3500.6; DB 2: Length 110000;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 3548; Conservative 0; Mismatches 24; Indels 3; Gaps 3;
QY 3342 ctactgcctacaagttaaggaaatccagaacttgaaacgaaacgagctctaccggccta 3401
Db CAACACGCCCTACAGAGTGAAGGAATCCAGAACTTGAACGAAACGACTCTACCCGGGCTA 77098
QY 3402 cactggactccaggatccagcgaattcttcattttcccgacagatataaccgctttt 3461
Db CACTGGACTGCCAGATCAGCGCAATTTGCAATTTTCCCGACAGATATACCCGCTTT 77158
QY 3462 catcagtgtgaagcagaagaagagactacttttaagtcaggagagagaggactcat 3521
Db CATCAGTGATGAAGCAGAGAGAGAGACTACTTTTAAGTCCAGGAGAGAGAGGGACTCAT 77218
QY 3522 tgcctagccagtcacctgggacctctgctcagagaccgcaccagagagctgcgccca 3581
Db TGCTCTGAGCCAGTCACTCGGACCTCTGCTCAGAGACCGCACAGAGGCTCGGCCCA 77278
QY 3582 ggatttgcgggacacgctgagtggaagcaggaaggaaggaagcagcagcagcagcagc 3641
Db GGATTGTTCGGGAGCCACGCTGAGTGGCAAGCAGGAGAGAGAGAGGAGGAGGAGGCTG 77338
QY 3642 accacagtgagagacagtgatgtggaaacacacagcgtgctcattcagcacctttgtt 3701
```

```
Db 77339 ACCACAGTGGAGGAGACAGAGTGGATGTGGAACACACAGGCTGCTCATTCAGCACCTTTGTT 77398
QY 3702 gttactgtgaacgtgaattgtggccagtatcaagagagctctctctgagtgactgcacat 3761
Db GTTACTGTGAACGTGAATGTGGCCAGTATCAAGAGAGTCTCTCTGAGTACTGCACCAT 77458
QY 3762 ggaactggcaccaggcgagactattagccaggcgagaccactagacttcagtcaggggacc 3821
Db GGCACGTGGCACCAGGGGGACTATTAGCCAGGGCAGACACACTAGACTTTCAGTGCAGGGACC 77518
QY 3822 tggtttcccttcgtttgcactttagtaaaattgggtggaggtttcccttttggactcttt 3881
Db TGGTTTTCCCTTCGTTTGCACCTTTAGTAAATTTGGTGGAGAGTTTCTCTTTGGATCTGTT 77578
QY 3882 ttgagactgttccagaagaaggcttcccttcccgagacacttccatagggcagcaatttg 3941
Db TTGAGACTGTTCAGAGAAAGAGGCTTCCTTTCCCGAGACACTTCATAGGCAGCAATTTTG 77638
QY 3942 gtgattcatttgcagaaaataactggctgtgttaattatttccctgccagcgccctgctg 4001
Db GTGATTCAATTTGCAGCAAAATACTGGCTTGTAAATTAATTTTCTGCCAGCGCTGCGTG 77698
QY 4002 ctaaacacagatgagatgagctaccactgaagctcgaagatgtgcctatttgaacgga 4061
Db CTAAACAACAGATCAGGATGAGCGTACCCTGAAGTCTGAAGATGTGCCATTGNACGGA 77758
QY 4062 cagtgatttccatagtcttcttaggtgttcttatgtcagatttccaaagccagcccccacag 4121
Db CAGTGTTTTTCATATGTTTCTAGTGTGCTTTATGTACAGTTTCCAAGCAGACCCCCACAG 77818
QY 4122 tgagaaatgttgagcaccgcacacacactgcaatgtgttttttaagtcagtgacac 4181
Db TGAGAAATGTGTAGGACCGCACACAACTGCAATGTGTTTTTAAGTCAAGGTGACAC 77878
QY 4182 atgtatttaagattttttttaaatactcttttcagtttaaatctcaactttttcaaacaa 4241
Db ATGTATTTAAGATTTTTTTTTTAAATCTCTTTGCGATTAAATCTCACATTTTTCACAAACAG 77938
QY 4242 cctgagatcagggaacaaacacttatttggcttttagctggagctcagcagggcagattg 4301
Db CCTGGATCAGGGCAAAACAACTTATATTTGTTTGTAGCTGAGGCTCAGCAGGAGATTTG 77998
QY 4302 caggcagggggcacttttcatccatgagcccgagcctcgggctggagctcgtgacac 4361
Db CAGGCAGGGGGCACTTTTCATCCATGAGGGCCCGCAGCCTGGGGCTGGGACTCTGATCAC 78058
QY 4362 cattgtgagggccagagcagctcgtatggagagaaatgtcaaaactgaacgaggttt 4421
Db CATTTGTGAGGCCAGAGGCGAGCTGCGTATGGAGAGAAATGTCAAACTGAACGAGGTTT 78118
QY 4422 caccactctaggaaagcagcttgttgagccctcagcctgagatgtgttagagggatggg 4481
Db CACCACCTTAGGAAAGAGCTTGTGTGAGCCCTCGACGTGGATGTGTAGAGGATGGG 78178
QY 4482 ctgaatgagcaggttagatttcctgcatacaacagtcttgggaagctgtgtgattcct 4541
Db CTGAATAGGCAGGTTAGATTTCCTGCATCAACAGTCTTTTGGGAAGCTGTGTGATTCCT 78238
QY 4542 gagaaacacagggagcgcagatggagccacacatgattgttcacccggtactgcagc 4601
Db GAGGAACAACAGGGAGCGGAGATGGAGCCACACATGAGTTTGTCTACCGGCTACTGCAGC 78298
QY 4602 acttgtaccagaatctcatgtccacaaaccccatgtlaaactttcaaccactcaaaagct 4661
Db ACTTTGTACCCAGAACTCTCATGTCCAAAACCCCATGTAAACTTTTCAACCACTCAAAAGCT 78358
QY 4662 gttatttcggctgaagaaataaacttttttctcaccagtcatttgtacctttcataat 4721
Db GTTTATTTCGGCTGAAGAAATAAATTTTTTCTCACCAGTCATTTGTACCTCTTCATAT 78418
QY 4722 ggcgtgtcgcacccctccagaaacgtgttatacttccagtcagtggtgggagagactgaag 4781
Db GGCTATGTGCGACCCCTCCAGAAACGTGTTATATCTCCAGTCAGTGTGGGAGAACTGAAG 78478
```

QY 4782 attcccggttggtcgaggaactgaggttgaccttcggaaggaagtcccaactatctta 4841
|||||
Db 78479 ACTTCCGGTTGGTCGAGAACTGAGGGTTGACCTTCGGGAGGAAGTTCACATCTATCTTA 78538
QY 4842 ttattatgctgtgtagtgggtcctgcaggggagacatccagtaactcgggtgtctttaat 4901
|||||
Db 78539 TTTATTATGCGTGTGATGTGGGTCTGCCAGGGAGACATCCAGTACTCGGTGTCTTTAAT 78598
QY 4902 tgcacactggggaactgtgtttattgacctcttttggggaactcctggttttggatgaagt 4961
|||||
Db 78599 TGCCACCTGGGAACTGTGTATTATGGCTTCTTTTGGGGATCCTGTGTTTGGATGAAGT 78658
QY 4962 gaggggaatacagaggtaaaagaattgtctccacctgaagcggggagtcgccgcttcaca 5021
|||||
Db 78659 GAGGGGAATACAGAGGTAAAGAAATTGTCTCCACCTTGAGCGGGGAGTCCCGCTTCACA 78718
QY 5022 ttctcgaaatggtgcagcactggggagacagttctgcgccggggcatggttgtttctcaa 5081
|||||
Db 78719 TTTCTGAAATGTTGTCAGCCACTGGGACAGTTCTGCCCGGGCATGGTTGTTTCTCAA 78778
QY 5082 ggtcctctaataataatccctattcttaataatccttggccctgatgggttttaagcaag 5141
|||||
Db 78779 GGTCTCTAATAATAATCCCTATTCTTTACATAATCCTTGCCCTTGATGGTTTTAAGCAAG 78838
QY 5142 aactcctgtgtcccatggtctccacctaccatccacctgctgttagcaagagtcctag 5201
|||||
Db 78839 AACTCTGTGTCCCATGGTCTCCACCACTCACCATCACCTGCTGTAGCAAGAGTCTCTAG 78898
QY 5202 tcagggaggtgcatatttagtagttaaatgcaacttaccatgagataaataaagagaga 5261
|||||
Db 78899 TCAGGGAGGTGCAATTTTAGTAGTTAAATTGCACTTATCCATGAGATAAATAAAGGAGA 78958
QY 5262 actgtttttatcagtggagtaacctaaatttcaagtgctgcctttttgaaattctg 5321
|||||
Db 78959 AGTGAATTTATCAGTGAGGCTAACTT-AAATTTACAGAGGTGCGCTTTTGTGAATCTTG 79017
QY 5322 ggcctctctctgtagaaccaaattggccctttgtggctcagcggcctgcacctaactgga 5381
|||||
Db 79018 GCCCTCTCTCTGTAGACCAATGGCCCTTGGTGGCTCAGCGGCTCGCACCTAACTGGA 79077
QY 5382 gagtctgagctcctgagctcaactgagcccacagactaggctcttggctccttcgc 5441
|||||
Db 79078 GAGTTCTGAGCTCTGTGAGCTCACTCAGCCACACAGACTAGGCTTCTTGGCTCTCTCCG 79137
QY 5442 agcatgctgtcactccccagaccgcgagctgtgggaagccatgtagggaggtctatt 5501
|||||
Db 79138 AGCATGGCTGCTACCCCCAGAACCCCGACGCTGTGGGAAGAGCCATGTAGGGAGGCTATT 79197
QY 5502 ccagggcatatacttccactgccttcagctgacgtcacagctgacaaaatcatctcctcta 5561
|||||
Db 79198 CCCAGGCATACACTTCCACTGCGCTTACGCTGACATCACAGTGCACAAATCATCTCCTCTA 79257
QY 5562 tcggagccagaagactcagctccacaaaatgaagtgtctgtcctgaaacaattcttgg 5621
|||||
Db 79258 TCGGAGCCACAAGACTTCAGCTCCACAAAATGAAGTGTCTGTCTCTGAAAACATTCCTTGG 79317
QY 5622 gaagaaatcccaactcgagaaaacgggtgtcctgtgaggttccacaagtctctgtttcat 5681
|||||
Db 79318 GAAGAATCCCAACATCGAGAAACGGTCTCTGTGAGTTTCCAAACATGCTTCTTGTTCAT 79377
QY 5682 ggggttcttcctgtatggagtggaatgaagagtggttttattttgttgttcttaactgagaaaa 5741
|||||
Db 79378 GGGTTTCTCCGTATGGAGTGGATTAAGAGTGTATTTATTTTGTGTCTTAAC TGAGAAAA 79437
QY 5742 aaaggaggaaccccaaggttgaggtcacacagtctccacagtttccagaggcgcttgg 5801
|||||
Db 79438 AAAGGAGGCCACCCAAAGTTGAGGTGCACAGTCTCCACAGTTTCCACAGTTTCCAGGAGCGTTTGG 79497
QY 5802 ggggtgggaagcacctcccagagcatgaggtctcaaggggacatgagtaagacatgtctg 5861
|||||
Db 79498 GGTGGGGAAGGCACCTCCAGAGCATGAGGCTCTAAGGGGACATGAGTAAGCATGTCTG 79557

QY 5862 tgaccagtgaggaagggagagccagctgcactcctgcagcggttctctagctgcagaa 5921
|||||
Db 79558 TGACCCAGTGAAGGAAGGAGAGCCAGCTGCACCTCTGCACGGGTTCCTAGTCGACAA 79617
QY 5922 gggtcocgccttagccagaggggaaaaacactgatagcagaagagccttggaatgcacacctg 5981
|||||
Db 79618 GGGTCCCGCCTTAGCCAGGGGAAACACCTGATAGCAGAAGAGGCTGGATGCACACCTG 79677
QY 5982 gcaagcgaagctctccgcgccagacacagtgctccatgtaagccctgcaactgggggtg 6041
|||||
Db 79678 GCACGCCGAGGCTCTCCGCCCAGACACAGTGTCCATGTACGCCCTGCACCTGGGGTGT 79737
QY 6042 gtgattcacgtgcacagatgccaaatcctgcacaaatccccacagatgggggaagtg 6101
|||||
Db 79738 GTGATTACGTGCACAGATGCCAATCTCTCACCAATATCCACAGATGGGGGAAGGTG 79797
QY 6102 agaggaaggggcaagtgtgttaactgtctcaagatgcttaaacctccatagagaga 6161
|||||
Db 79798 AGAGGAAGGGCAAGTGTAACTGCTCAAGAGATGCTTAAACCTCCATAGAGAGGA 79857
QY 6162 gccgggcgcaggggcatctgtgtcccgtaacacactgcagcagggaaggtgctggc 6221
|||||
Db 79858 GCCGGGCGCAGGGGCATCTGTGTCCGTCACACACTGCAGCAGGGAAGGTGGCTGGC 79917
QY 6222 tggctcctggcatcagtggtttgttttaagctccagaggtcttattgccaattgtcttt 6281
|||||
Db 79918 TGGCTCCTCGCATCAGTGGTTTGGTTTAAGCTCCAGAGGCTCTTATTGCCATTGTCTTT 79977
QY 6282 tcctctgccccttgagccagcctaaggccctggagctgttcttctttaggcggaatgaactg 6341
|||||
Db 79978 TCCCTGTGCCCTTGAGCCAGCCTAAGGCCCTGGAGTCTGTCTTCTTAGCGCGATGAAC TG 80037
QY 6342 acatgctcctaccatgaccaggtctctgggcaagctcctcaactcctcaactggaaggtgg 6401
|||||
Db 80038 ACATGCTCCTACCATGACAGGCTCTGGGCAAGGCTCCTCACAGTATCCTTGTAGAGGTGG 80097
QY 6402 gcatggaggtgcccatttctcaggtacagaaaccttcagagagataaataagcttgcct 6461
|||||
Db 80098 GCATGGAAGTGCCCATTTCTCAGGTACAGAAACCTTCAGAGAGGATAAATAGCTGCCCT 80157
QY 6462 gtagaagcaggactgaaaaccttgcggcctgactccccagctactctgcccactgtag 6521
|||||
Db 80158 GTAGAAGCAGGACTGAAACCCCTTGTCCGCTGACTTCCCCAGCTACTCTGCCCACTGTAG 80217
QY 6522 cccctgcttactgtctgagcacccctcaccatcctgtataccttaaatatacaaga 6581
|||||
Db 80218 CCCCTGCCCTTACTGTCTGGCACACCCCTCACCATCCTGTATACCTTAAATATCAAGA 80277
QY 6582 gggcaagagagaaagggctttaaagataagttattttttaaggaaaccttaattattt 6641
|||||
Db 80278 GGGCAAGAGAGAAAGGGCTTTAAAGATAAGTTATTTTTTAAAGAACCTTAATATTATT 80337
QY 6642 ttaagaagttaacaaaaatagtgacgtgaaatgc-aaaaaataaaaaaataaactgact 6700
|||||
Db 80338 TTAAGAAGTAAACAAATAGTGTAGCTGAAATGCAAAAAAATAAAAAAATCTGACT 80397
QY 6701 acccttttgaatactgtcttccagattgtttttatatgttaattcttagacactgttcat 6760
|||||
Db 80398 ACCCTTTGAAATGCTGCTTCAGATTGTTTTTATATGTAATTTCTTAGACACTTGTCTAT 80457
QY 6761 taagaaaaatagtggtggtggttctcagcaagaagcacactgcgcagtggtggttggat 6820
|||||
Db 80458 TAAGAAAAATAGTGGCTGGCTGTGCTCAGCAAGAAGCACACTGCGAGCTGGCTTTGGTAT 80517
QY 6821 aggaagtgggaagcaagacctgggtttctgacaagtgcgctgaagacttaaccttccatc 6880
|||||
Db 80518 AGGAAGTGGGAAGGCAAGGACCTGGGTTTCTGACAAGTGCAGTGCAGATACACATCCATC 80577
QY 6881 tggagagctggtggttctgtccctctgggtagggc 6915
|||||
Db 80578 T-GAGAGCTGCTTGCCTTGTTCCTCGTGGTAAGGCC 80611

RESULT 5

BC004539 1462 bp mRNA linear PRI 12-JUL-2001
 Homo sapiens, clone IMAGE:3948153, mRNA, partial cds.

DEFINITION
 BC004539
 BC004539.1 GI:13528680

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 1462)

Strausberg, R.

Direct Submission

Submitted (14-MAR-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs-r@mail.nih.gov

Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Rubin Laboratory

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc.mgc@nih.gov

Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,

Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,

Dietrich, N.L., Guan, J., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,

Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,

Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,

Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 15 Row: d Column: 11

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

FEATURES

source

Location/Qualifiers

1..1462

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3948153"

/tissue_type="Ovary, adenocarcinoma"

/clone_lib="NIH_MGC_9"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

<1..739

/codon_start=2

/product="Unknown (protein for IMAGE:3948153)"

/protein_id="AAH04539.1"

/db_xref="GI:13528681"

/translation="ARGLFDLADRMQKVNCKSSAEVQCLLSQRRIFRAGSLCKRK

SPCKDXTICTDLGVALCCQKSGYFQNKMDHSCRACEDGYRLENETCMSCPFGLG

GLNCGNPYLITVIAAGGGLLLILGIALIVTCRKRNKNDISKLIKFGDFQMSFYA

EYKPNPSREWREATEMHENGSTKNLLQMTDVIYVSVNPELERNGLYPAYTGLP

GSRHSCIFFGQYNPSFISDESRRDYF"

BASE COUNT 384 a 342 c 380 g 356 t

ORIGIN

Query Match 20.7%; Score 1438.8; DB 9; Length 1462;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1440; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2769 gctattgacctggctgataggatgcagaaatgtgtcaactcctgcgaagtcctctgcta 2828

|||||

Db 10 GCTATTGACCTGGCTGATAGGATGCAGAAATGTGTCAACTCCTGCAAGTCTCTGCTGA 69

QY 2829 ggtctgcagctcttgggatactcagagggcgagatctttagagcgggcagcttgttcaagcg 2888
 |||||
 Db 70 GGTCTGCAGCTCTTGGGATCTCAGAGCGGATCTTTAGAGCGGCGAGCTTTGTGCAAGCG 129
 |||||
 QY 2889 gaagagtcctccgaatgtgacaaagacacctccatctgcaactgacctggagcggcttgccct 2948
 |||||
 Db 130 GAAGAGTCCCAGTGTGACAAAGACACCTCCATCTGCACCTGACCTGGACGGCTTGCCCT 189
 |||||
 QY 2949 gtgcagtgcaagtcgggatacttccagtttcaacaagaatggaccactcctccgagcatg 3008
 |||||
 Db 190 GTGCCAGTGCAGTCCGGGATACTTTTCAAGTCAACAAGATGACCACTCTCTGCCAGCATG 249
 |||||
 QY 3009 tgaagatggaatagcttcaaaatgaacctgcaatgaactgagttgcccatttgcccttggtg 3068
 |||||
 Db 250 TGAAGATGGATATAGGCTTGAAGATGAACCTGCATGATGTGCCCATTTGGCTTTGGTGG 309
 |||||
 QY 3069 tctcaactgtggaacccctcatcctcagcttctcactgtgtgctcagcgcggggaggtgg 3128
 |||||
 Db 310 TCTCAACTGTGGAACCCCTATCAGCTTATCAGTGTGTGTGATCGCAGCGCGGAGGTGG 369
 |||||
 QY 3129 gctcctgctcaactagcgcagctgctgctgctgctgctgctgctgctgctgctgctgctg 3188
 |||||
 Db 370 GCTCCTGCTCATCTTAGCATCGCACTGATTTTACCTTTGCAGAAAGATAAATAATGA 429
 |||||
 QY 3189 cataagcaaacctcatcttcaaaagtggagatttccaaatgtcccccatatgctgaatccc 3248
 |||||
 Db 430 CATGAAGCAAACTCATCTTCAAAAGTGGAGATTTCCAATGTCCCGGTATGCTGAATACCC 489
 |||||
 QY 3249 caaaatcctcgtctcaaaagaatggggccgagaagctattgaaatgcatgagatggag 3308
 |||||
 Db 490 CAAAAATCCTCGCTCACAAGAATGGGCGGAGAGCTATTGAAATGCATGAGATGGAAG 549
 |||||
 QY 3309 taccaaaacctcctcagatgacgagtgctgactactcgcctcagatgtaagaatccc 3368
 |||||
 Db 550 TACCAAAACCTCCTCAGATGACGGATGTGTACTACTCGCTCAAGTGTAAAGGAATCC 609
 |||||
 QY 3369 agaactgaaacgaacggactctaccggcctacactgactgactgactgactgactgactg 3428
 |||||
 Db 610 AGAAGTGAAGCAAAACGGAGCTCTACCGGGCTACACTGGAGTGCAGGATCACGGCATTC 669
 |||||
 QY 3429 ttgcatttccccggagcagatataaacccgtctttcatcagtgatgaaagcagaagaaga 3488
 |||||
 Db 670 TTGCATTTTCCCGGACAGATATAACCCGCTTTTCATCATGATGATGAAAGCAGAGAAGA 729
 |||||
 QY 3489 ctacttttaagtcag 3548
 |||||
 Db 730 CTACTTTTAAAGTCCAGGAG 789
 |||||
 QY 3549 tgctcag 3608
 |||||
 Db 790 TGCTCAGAGGAG 849
 |||||
 QY 3609 caagcagaag 3668
 |||||
 Db 850 CAAGCAG 909
 |||||
 QY 3669 ggaacacagagctgctcattcagcagcctttgttacttgaacgtaagtgtgggccaag 3728
 |||||
 Db 910 GGAACACAGAGCTGCTCATTCAGCAGCCTTTGTTGTTACTGTGAACGTAATGTGGGCCAG 969
 |||||
 QY 3729 tatcaagagagatcctctgagtgactgcacatggcactggcagcagcagcagcagcagc 3788
 |||||
 Db 970 TATCAAGAGAGATCTCTCTGAGTGACTGCACCATGGCACTGGCACAGGGCGACATATTAGC 1029
 |||||
 QY 3789 cagggcagacacactagacttcagtgagggagagagagagagagagagagagagagagag 3848
 |||||
 Db 1030 CAGGGCAGACACTAGACTTCAGTGCAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1089
 |||||
 QY 3849 aaattgggtggaggttctcttttgatctgttttgagactgttccagaaagaaggttc 3908
 |||||
 Db 1090 AAATTGGGTGGAGGTTTCTCTTTTGGATCTGTTTGGAGACTGTTCAGAAAGAGGCTTC 1149
 |||||

Qy	3909	cttcccgagacacttccatgagcagcaatttggatctatcttbgcagcaaaatactggc	3968
Db	1150	CTTTCCCGAGACACTTCCATAGGAGCAAAATTGGTGATTCATTTGCGACAAAATACTGGC	1209
Qy	3969	ttgttaattatttcttcgcccagcgcctgctgctaaacaacacagatgagcgctac	4028
Db	1210	TTGTGTAATATTTTCCCTGCCCAGGCCCTGCGTGCTAAACAACAGATGAGGATGACGCTAC	1269
Qy	4029	cactgaagctcgaagatgtcgccattgaacggcagcagtgcttttcataatgtttctaggttgt	4088
Db	1270	CACGGAAGTCTGAAGATGTCGCCATTGAACGACAGATGCTTTTCATATGTTTTCTAGGTGT	1329
Qy	4089	cttatgtctacagtttccaaagccagcccccacagtgaggaaatgtgtgagccacgcacac	4148
Db	1330	CTTATGTCTACAGTTTCCAAAGCCAGCCGCCACAGTGAGGAAATGTGTGAGGCACGCCACAC	1389
Qy	4149	aactgcaatgtgtttttaaagtcgaagtgacacatgtatttaagaatttttttttaaaatc	4208
Db	1390	AACGTCAATGTGTTTTTAAAGTCAAGGTGACACATGATTATTAAGATTTTTTTTTTAAATC	1449
Qy	4209	tc 4210	
Db	1450	TC 1451	
RESULT	6		
LOCUS	BC014053	1397 bp mRNA linear	PRI 04-DEC-2001
DEFINITION	Homo sapiens, clone MGC:20272 IMAGE:3632959, mRNA, complete cds.		
ACCESSION	BC014053		
VERSION	BC014053.1	GI:17315133	
KEYWORDS	MGC.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 1397)		
TITLE	Strausberg, R.		
JOURNAL	Direct Submission		
	Submitted (10-SEP-2001), National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/nisc_mgcengr1.nih.gov Contact: nisc_mgcengr1.nih.gov Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripo, S., Thomas, P.J., Tlonsong, E.E., Touchman, J.W., Tsourgeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.		
FEATURES	Location/Qualifiers		
source	1..1397 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="MGC:20272 IMAGE:3632959"		


```

Db 789 ACAGTGGGGAAGGTGAGAGGAGGCGCAAGTGATGTAACTGCTCAAGAGATGCTTA 848
QY 6145 aacctccatagagagcgagcgagggcagcctctgtgtcccgccacacactgcagc 6204
Db 849 AACCTCCATAGAGGAGGCGGCGGCGAGGCGACTCTGTGTCTCCCGTCACACATCGCAGC 908
QY 6205 aggaaggggtggctggctccctggcctgcagtcagtggtttggttaagctccagagggtc 6264
Db 909 AGGGAAGGGTGGCTGGCTGCTCCCTGGCATCAGTGGTGGTTTAAAGCTCCAGAGGTTTC 968
QY 6265 ttattgcattgttttctctctgccttgagccagccttaagggcccttgagctgttttc 6324
Db 969 TTATTGGCATGTCTTTTCTCTCTGCGCTTGGAGCGACGCTTAAGGCGCTGGAGTCTGTTC 1028
QY 6325 tttagcgagatgaactacatgctctaccatgacacagcctctgggcaagcctctccaca 6384
Db 1029 TTTAGGCGGATGAACATGACATGCTCTACCATGACCAAGGCTCTGGGCAAGGCTCTCACA 1088
QY 6385 gtatccttgagtggtggcagtggaagtgcacattctcaggtacagaaaccttcagagag 6444
Db 1089 GTATCCTTGAGAGTGGGCGATGGAAGTGCCATTTCTCAGGTACAGAAACCTTCAGAGAG 1148
QY 6445 gataaatagcttgccctgtagaagcagtgactgaacaccttgcgcctgactccccagc 6504
Db 1149 GATAAATAGCTTGCCCTGTAGAAGCAGGACTGAAACCCCTTGTTCGGCCTGACTCCCCAGC 1208
QY 6505 tactctccactgtagccctcttactgtctctgagcaccctccaccatcttat 6564
Db 1209 TACTCTCCCACTGATGCCCTGCTTACTGTCTCTGGCACACCCCTCACCATCTCTAT 1268
QY 6565 accttaatatcaagagggcagagagaaaggctttaaagataagttatctttttaaag 6624
Db 1269 ACCTTAATATCAAGAGGCGCAGAGAGAAAGGCTTTAAAGATAAGTATTTTAAAG 1328
QY 6625 gaaccttaattatttttaagaagtagtaaccaaatagtcagtgaaatgcaaaaaaaa 6684
Db 1329 GAACCTTAATATTATTTTAAAGAGTAGTAACCAAAATTAGTGACGTGAATGCAAAAAA 1388
QY 6685 aaaaaaaa 6692
Db 1389 AAAAAAAA 1396

RESULT 7
AC092983_0/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AC092983 Accession AC092983
Fragment Name Begin End
AC092983_0 1 110000
AC092983_1 100001 210000
AC092983_2 200001 310000
AC092983_3 300001 350921
LOCUS AC092983 350921 bp DNA linear HTG 12-JAN-2002
DEFINITION Homo sapiens chromosome 3q clone RP11-717C3, WORKING DRAFT
SEQUENCE, 49 unordered pieces.
ACCESSION AC092983.2 GI:18139248
VERSION AC092983
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 350921)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratinge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

```

```

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaracne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovari,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseghe,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogulu,H., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaik,J., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL
REFERENCE
2 (bases 1 to 350921)
AUTHORS
Worley,K.C.
TITLE
JOURNAL
Direct Submission
Unpublished
2 (bases 1 to 350921)
Worley,K.C.
Direct Submission
Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 12, 2002 this sequence version replaced gi:15136886.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HDEZ
Center clone name: RP11-717C3
----- Summary Statistics
Assembly program: Phrap; version 0.990329first call to
findPhrapList
Consensus quality: 333925 bases at least Q40
Consensus quality: 347288 bases at least Q30
Consensus quality: 355437 bases at least Q20
Estimated insert size: 353663; sum-of-ctngs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.6x in Q20 bases; sum-of-ctngs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
consists of 49 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 50319: contig of 50319 bp in length
* 50419: gap of unknown length
* 50420: contig of 21913 bp in length
* 72332: gap of unknown length
* 72333: gap of unknown length
* 72433: contig of 22122 bp in length
* 94555: gap of unknown length

```



```

Db 41063 TGCCCCACTTTCTGTCTCACAACAACTTGCCACAGTCATCTTCTACCCCTGTCTGCTGCC 41004
Qy 1596 cagggaagagagactcctgtgacttcattcagacatcaacaatgacatcattcatgac 1655
Db 41003 CAGGCAAGGAGAGACTCCTGTGACTTCAATTTTCAGACATCAACAATGACATCATTCATGAC 40944
Qy 1656 aatgtccatagtagtcaaaactgcagaccttaagagccagagcagacccacacacagagaa 1715
Db 40943 AATGCTCCATAGTAGTCAAACTGCAGACTTAAAGAGCCAGAGACCCCCACACCAAGAGAA 40884
Qy 1716 agtcattacagaataaagtcaccaagcctggtgtctctgtccacagagtcacccaagc 1775
Db 40883 AGTCATTACAGAAATCAAAAGTCACCAAGCCTGGTGTCTCTGCTGCCACAGAGTCCACCAAGC 40824
Qy 1776 tgtacaacaactcctcttgctccatcttaacagagtcctccacagagcaaacct 1835
Db 40823 TGTAAACAACAATCTCTCTTGGCTCCATCTTAAACAGAGTCTCCACAGAGCAAAACCT 40764
Qy 1836 tccagccaaagcaccacttagcacaatgtctcacaatttcacaattcaccattctgaa 1895
Db 40763 TCCAGCCACAGCACCACCTTAGCACAATGTCTCCAATTTCACAACTACCATCTCTGAA 40704
Qy 1896 gacctctcagcctcttatgacctctctgacctgtcaagcagcagcatctctggtcac 1955
Db 40703 GACCTCTCAGCCTCTTATGACCACCTCTGCGCACCCCTGTCAAGCAGCAGCATCTCTGGTCAC 40644
Qy 1956 tggccctatagcctacagactacagcttggaacagctctcgtgacccatcctgaaat 2015
Db 40643 TGGCCCTATAGCCGTACAGACTACAGCTGGAACACAGCTCTCGTGCACCCATCTCTGAAAT 40584
Qy 2016 actggtctcctcaaatctcaacagaagtggtgcatcagcagacaagaacccagtgattgt 2075
Db 40583 ACTAGTTCTCAATCTCAACAGAGGTGGCATCAGCAGACAAAGAACCCAGTGATTGT 40524
Qy 2076 ggatgtaccactggattgattccctttgaccagtgatccaccatcagcaaaagaatgac 2135
Db 40523 GGATGCTACCACTGGATTGATCCCTTTGACCAGTGTACCCACATCAGCAAAAGAAATGAC 40464
Qy 2136 cacaagcttgcttacagcagatagacccagcttcacagttcccttcggaacatctcc 2195
Db 40463 CACAAGCTTTGGCGTTACAGCAGAGTACAGCCCAAGCTTCAGTTCCTTCGGAACATCTCC 40404
Qy 2196 ttctcccaaacacacagttgtttccagcgtgaagacttggtcccaaatctgccaacct 2255
Db 40403 TTCTCCCAAAACCACAGTTGTTTCCACGGCTGAAGACTTGGCTCCCAATCTGCCACCTT 40344
Qy 2256 tgcgttcagagcagcacacagtcacacacacactgctccttccttcagcctcagtcacaa 2314
Db 40343 TGCTGTTTCAGAGCAGCACACAGTACCAACAACAGTGTCTCTTCAGCCTTCAGGTAAAA 40285

RESULT 8
AC026629 AC026629 118407 bp DNA linear HTG 04-JUN-2000
LOCUS Homo sapiens chromosome 3 clone RP11-816I3 map 3, LOW-PASS SEQUENCE
DEFINITION
ACCESSION AC026629
VERSION AC026629.2 GI:8247896
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 118407)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,

```

```

Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Hearford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 4, 2000 this sequence version replaced gi:7284654.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8653
Center clone name: 816_I_3
-----
* NOTE: this record contains 141 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 757: contig of 757 bp in length
* 758 857: gap of 100 bp
* 838 1584: contig of 727 bp in length
* 1585 1684: gap of 100 bp
* 1685 2449: contig of 765 bp in length
* 2450 2549: gap of 100 bp
* 2550 3306: contig of 757 bp in length
* 3307 3406: gap of 100 bp
* 3407 4163: contig of 757 bp in length
* 4164 4263: gap of 100 bp
* 4264 5024: contig of 761 bp in length
* 5025 5124: gap of 100 bp
* 5125 5878: contig of 754 bp in length
* 5879 5978: gap of 100 bp
* 5979 6734: contig of 756 bp in length
* 6735 6834: gap of 100 bp
* 6835 7583: contig of 749 bp in length
* 7584 7683: gap of 100 bp
* 7684 8446: contig of 763 bp in length
* 8447 8546: gap of 100 bp
* 8547 9304: contig of 758 bp in length
* 9305 9404: gap of 100 bp
* 9405 10161: contig of 757 bp in length
* 10162 10261: gap of 100 bp
* 10262 11003: contig of 742 bp in length
* 11004 11103: gap of 100 bp
* 11104 11831: contig of 728 bp in length
* 11832 11931: gap of 100 bp
* 11932 12694: contig of 763 bp in length
* 12695 12794: gap of 100 bp

```

TITLE
JOURNAL

COMMENT

Db 84482 ACAATGCTTCTGTTTCATGGGTTTCTTCCTGATGGAGTGAAGAGTGTATTTTATTTG 84541
 QY 5724 ttgttctaactgagaaaaaagaggagcaccacaaagtgtgaggtcacacagctccacag 5783
 Db 84542 TTGTTCTTAAGTGAAGAAAAAGAGGAGCCACCAAGGTTGAGGTTCACACAGTCTCCACAG 84601
 QY 5784 ttccagagagcgtttggggtgggaggaagcactccagagcatgagctcgaaggagac 5843
 Db 84602 TTTCCAGAGGCGTTTGGGGTGGGGAAGGACCTCCAGAGCATGAGGCTTAAGGGGAC 84661
 QY 5844 atgagtaaacatgtctgtgacccagtgaggaagggagagggcagctgcactcctgcacy 5903
 Db 84662 ATGAGTAAGCATGCTGTGTGACCCANTGAGAGGGAGAGGCCAGCTGCATCTCTGCACG 84721
 QY 5904 gggttctaactgagaaaggtcccgctaggccaggaaggaacacatgatagcagaaga 5963
 Db 84722 GGGTTCCTAGCTGCAGAAAGGTCCTCGGCTTANGCGAGGGGAAACACCTGTATGACGAAG 84781
 QY 5964 ggctggtatgcacacctggcagcgcaggtctctcgcccgagacacagctgctccatgtcag 6023
 Db 84782 AGCTGGATGCACACCTGGCAGCGGAGGCTCTNGCGCCAGACACAGCTGTNNNNNNN 84841
 QY 6024 ccctgcacctggggtgtgtattcagctgacagatgacacaaatcctgcaccaatctcc 6083
 Db 84842 NNN 84901
 QY 6084 cacagatgggggaaggtgagagaaggggcaag 6116
 Db 84902 NNN 84934

RESULT 9
 AC026629/c
 LOCUS
 DEFINITION Homo sapiens chromosome 3 clone RP11-81613 map 3, LOW-PASS SEQUENCE
 SAMPLING.
 ACCESSION AC026629
 VERSION AC026629.2 GI:8247896
 KEYWORDS HTG: HTGS_PHASE0.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 118407)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 3, clone RP11-81613
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 118407)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bieda,F.,
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
 Dodgson,J., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Gargan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienda,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Jun 4, 2000 this sequence version replaced gi:7284654.

All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L8653
 Center clone name: 816_I_3

 * NOTE: This record contains 141 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 757: contig of 757 bp in length
 * 758 857: gap of 100 bp
 * 858 1584: contig of 727 bp in length
 * 1585 1684: gap of 100 bp
 * 1685 2449: contig of 765 bp in length
 * 2450 2549: gap of 100 bp
 * 2550 3306: contig of 757 bp in length
 * 3307 3406: gap of 100 bp
 * 3407 4163: contig of 757 bp in length
 * 4164 4263: gap of 100 bp
 * 4264 5024: contig of 761 bp in length
 * 5025 5124: gap of 100 bp
 * 5125 5878: contig of 754 bp in length
 * 5879 5978: gap of 100 bp
 * 5979 6734: contig of 756 bp in length
 * 6735 6834: gap of 100 bp
 * 6835 7583: contig of 749 bp in length
 * 7584 7683: gap of 100 bp
 * 7684 8446: contig of 763 bp in length
 * 8447 8546: gap of 100 bp
 * 8547 9304: contig of 758 bp in length
 * 9305 9404: gap of 100 bp
 * 9405 10161: contig of 757 bp in length
 * 10162 10261: gap of 100 bp
 * 10262 11003: contig of 742 bp in length
 * 11004 11103: gap of 100 bp
 * 11104 11831: contig of 728 bp in length
 * 11832 11931: gap of 100 bp
 * 11932 12694: contig of 763 bp in length
 * 12695 12794: gap of 100 bp
 * 12795 13547: contig of 753 bp in length
 * 13548 13647: gap of 100 bp
 * 13648 14410: contig of 763 bp in length
 * 14411 14510: gap of 100 bp
 * 14511 15288: contig of 758 bp in length
 * 15289 15388: gap of 100 bp
 * 15389 16119: contig of 751 bp in length
 * 16120 16219: gap of 100 bp
 * 16220 16986: contig of 767 bp in length
 * 16987 17086: gap of 100 bp
 * 17087 17816: contig of 730 bp in length
 * 17817 17916: gap of 100 bp
 * 17917 18663: contig of 747 bp in length
 * 18664 18763: gap of 100 bp
 * 18764 19505: contig of 742 bp in length
 * 19506 19605: gap of 100 bp
 * 19606 20362: contig of 757 bp in length
 * 20363 20462: gap of 100 bp
 * 20463 21180: contig of 718 bp in length
 * 21181 21280: gap of 100 bp


```
VERSION      AX245407.1  GI:15860081
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 400)
AUTHORS      Williams L.T., Escobedo J., Innis M.A., Garcia P.D.,
             Sudduth-Klinger J., Reinhard C., Randazzo F., Kennedy G.C., Pot D.,
             Kassam A., Lamson G., Drmanac R., Crkvenjakov R., Dickson M.,
             Drmanac S., Labat I., Leshkowitz D., Kita D., Garcia V. and
             Stache-Crain B.
TITLE        Human genes and gene expression products
JOURNAL      Patent: WO 0166753-A 337 13-SEP-2001;
             Chiron Corporation (US); Hyseq Inc. (US)
FEATURES     Location/Qualifiers
             source
               1..400
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
BASE COUNT   97 a 93 c 123 g 87 t
ORIGIN
Query Match      5.6%; Score 387.4; DB 6; Length 400;
Best Local Similarity 99.7%; Pred. No. 2.1e-90;
Matches 388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5617 cttgggaagaatcccaacatcagaaacgggtgctctgtgagttccaaacatgctcttg 5676
Db 12 CTTGGGAAGATCCCAACATCAGAAACGGGTGCTCTGTGAGTTCCAAACATGCTTCTTG 71

QY 5677 ttcattggtttccgtatgagtgattaaagagttttattttgttcttaactga 5736
Db 72 TTCATFGGTTTCTCCGTATGAGATGGATTAAGAGTGTATTATTGTGTCTTAACATGA 131

QY 5737 gaaaaaaggaggcaccacacaggttgaggttcacacagttccacagttccacaggagcg 5796
Db 132 GAAAAAAGGAGGACCACACAGGTGAGGTTCACACAGTCTCCACAGTTTCCAGGAGCG 191

QY 5797 tttgggggtgggaagcaccctcagagcatgaggtctcttaagggtgacatgagtaagcat 5856
Db 192 TTTGGGGGTGGGAAGCACCCTCAGAGCATGAGGCTCTAAGGGGACATGAGTAAGCAT 251

QY 5857 gtctgtaccagtgagaaaggagagcagctgcactctctgcaggggttcctagctg 5916
Db 252 GTCTGTACCAGTGAAGGAGGAGAGCCAGCTGCACCTCTCTGACGGGGTTCCTAGCTG 311

QY 5917 cagaaggttcgcctcctaggccaggaggaacacacacctgatagcagaagggcctggatcac 5976
Db 312 CAGAAGGTTCGCCTATGCGAGGGGAAACACCTGATAGCAGAAGAGGCCCTGGATGCAC 371

QY 5977 acctggcacgcagggtctctccgccaga 6005
Db 372 ACCTGGCACGCAGGCTCTCCGCCAGA 400

RESULT 11
AC092983_3
WPCOMMENT
Sequence split into 4 fragments LOCUS AC092983 Accession AC092983
Fragment Name Begin End
AC092983_0 1 110000
AC092983_1 100001 210000
AC092983_2 200001 310000
AC092983_3 300001 350921
Continuation (4 of 4) of AC092983 from base 300001 (AC092983 Homo sapiens chromosome 3q

Query Match      4.7%; Score 326.2; DB 2; Length 50921;
Best Local Similarity 92.5%; Pred. No. 4.1e-74;
Matches 343; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 606 ctttgggaacatcagcttccagcagctctgaggtgcacaaatgaagtcccatgtctca 665
```

```
Db 31390 CCTCGGAACATCAGCTTGCAGCAGCTCTGAGGTGCAAAATGAACTCCCATCTCTCA 31449
QY 666 gactgagactgtctagtgactgcacccatgagagtgagagatcactgcacactg 725
Db 31450 GACTGAGACTGTCTAGTGGTCGCACCCATGAGAGTGAGAGATCAGCTGTACACTG 31509
QY 726 gctttgaccacagacacacatctgcagatgtgcagagaagctctgttctatctctga 785
Db 31510 GCTTTGACCACAGCACAACATCTGCAGATGTGCAGAGAGCTCTGTCTTATATCTGA 31569
QY 786 aggtgtgaactctcagttgaccacagttctcagactctactgtacagttctggaggaag 845
Db 31570 AGGTGTGAATGCTTCAGTGTGACCCAGTTCACAGCTCTACTGTACAGTCTGGAGGAAG 31629
QY 846 tcacacagcattgggagatagagttattcagagttcttcacatctctcttcctcgaaag 905
Db 31630 TCACACAGCATTTGGGAGATAGGAGTTATTTCAGAGTCTTCATCTACATCTTCTCGGAAAG 31689
QY 906 cttgaattcatcagcaccacagctggagaagcttcaatcgtctgggattagctacaggtcaagt 965
Db 31690 CTTGAATTCATCAGCACCACCGTGGAGACGTTTCGAGTGAGTTCCTTCATTCATCAACT 31749
QY 966 gctgtggcacag 976
Db 31750 GTGCCCCATAG 31760

RESULT 12
AC069571
LOCUS
DEFINITION Homo sapiens chromosome 3 clone CTD-2015F14 map 3, WORKING DRAFT
SEQUENCE, 16 unordered pieces.
AC069571
ACCESSION AC069571.3 GI:10140818
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 115766)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE        Homo sapiens chromosome 3, clone CTD-2015F14.
JOURNAL
REFERENCE    2 (bases 1 to 115766)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,E.,
             Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
             Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
             Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Gage,D.,
             Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
             Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
             Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
             Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
             Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
             Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
             McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
             Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
             Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
             O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
             Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
             Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
             Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
             Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
             Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J.,
             Young,G., Zainoun,J., Zimmer,A. and Zody,M.
             Direct Submission
             Submitted (03-JUN-2000) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
             On Sep 15, 2000 this sequence version replaced gi:9887750.
             All repeats were identified using RepeatMasker:
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
```


----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8478
Center clone name: 2015_F_14
----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 106109 bases at least Q40
Consensus quality: 111659 bases at least Q30
Consensus quality: 113401 bases at least Q20
Insert size: 121000; agarose-fp
Insert size: 114266; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 17190: contig of 17190 bp in length
* 17191 17290: gap of 100 bp
* 17291 18431: contig of 1141 bp in length
* 18432 18531: gap of 100 bp
* 18532 20008: contig of 1477 bp in length
* 20009 20108: gap of 100 bp
* 20109 22046: contig of 1938 bp in length
* 22047 22146: gap of 100 bp
* 22147 24142: contig of 1996 bp in length
* 24143 24242: gap of 100 bp
* 24243 27247: contig of 3005 bp in length
* 27248 27347: gap of 100 bp
* 27348 31214: contig of 3867 bp in length
* 31215 31314: gap of 100 bp
* 31315 36549: contig of 5235 bp in length
* 36550 36649: gap of 100 bp
* 36650 42234: contig of 5585 bp in length
* 42235 42334: gap of 100 bp
* 42335 45924: contig of 3590 bp in length
* 45925 46024: gap of 100 bp
* 46025 50925: contig of 4901 bp in length
* 50926 51025: gap of 100 bp
* 51026 59799: contig of 8774 bp in length
* 59800 59899: gap of 100 bp
* 59900 83502: contig of 23603 bp in length
* 83503 83602: gap of 100 bp
* 83603 92991: contig of 9389 bp in length
* 92992 93091: gap of 100 bp
* 93092 103204: contig of 10113 bp in length
* 103205 103304: gap of 100 bp
* 103305 115766: contig of 12462 bp in length.

FEATURES

Source

1. .115766
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="CTD-2015F14"
/clone_lib="CIRD Human BAC"
1. .17190
/note="assembly_fragment
clone_end:SP6
vector_side:left"
17291. .18431
/note="assembly_fragment"

misc_feature

misc_feature

misc_feature 18532. .20008
/note="assembly_fragment"
misc_feature 20109. .22046
/note="assembly_fragment"
misc_feature 22147. .24142
/note="assembly_fragment"
misc_feature 24243. .27247
/note="assembly_fragment"
misc_feature 27348. .31214
/note="assembly_fragment"
misc_feature 31315. .36549
/note="assembly_fragment"
misc_feature 36650. .42234
/note="assembly_fragment"
misc_feature 42335. .45924
/note="assembly_fragment"
misc_feature 46025. .50925
/note="assembly_fragment"
misc_feature 51026. .59799
/note="assembly_fragment"
misc_feature 59900. .83502
/note="assembly_fragment"
misc_feature 83603. .92991
/note="assembly_fragment"
misc_feature 93092. .103204
/note="assembly_fragment"
misc_feature 103305. .115766
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 31539 a 24759 c 26611 g 31352 t 1505 others
ORIGIN

Query Match 3.8%; Score 266.4; DB 2; Length 115766;
Best Local Similarity 99.6%; Pred. No. 2.4e-56;
Matches 267; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gttcgatgaagaattgcgcgttttcaacaagaagtgaacagctcgagatgggaac 60
|||||
Db 110015 GTTCGATGAAGAATTGCCGCTTTTCAACAAGAAGTGAACAGCTCGAGATGGGAAC 110074

Qy 61 agagaggcgatggggctgtcagaagaatggactgtgcacagccaaggagccacttc 120
|||||
Db 110075 AGAGAGGCGGATGGGGCTGTTCAGAAGAATGGACTGTGCACAGCCAAGAGGCCACACATTC 110134

Qy 121 gcttgagcccttccttctctctgttggatggagctgacacagccttctag 180
|||||
Db 110135 GGCTGGAGCCCTTCCTTCTCTCTGCTTTGGAGATGGGAGCTGACACGCCCTTCTAG 110194

Qy 181 gaagagaaattcctcagaccagatctctcctggctgatttctacagacagcgttc 240
|||||
Db 110195 GAAGAGAAATTCCTTCAGGACCAGATCTCTCTGGCTGCATTTCTACAGGACAGCAGCTTC 110254

Qy 241 ctctctctcttagaccttctcactc 268
|||||
Db 110255 CTCCTCTCTTAGACCTTTCTCTCACGT 110282

RESULT 13

AF389852

LOCUS AF389852 34308 bp DNA linear PRI 23-AUG-2001
DEFINITION Homo sapiens psoriasis susceptibility gene candidate interval,
partial sequence.
ACCESSION AF389852
VERSION AF389852.1 GI:15281554
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 34308)
AUTHORS Hewett,D.R.

	Query Match	1.88;	Score 126.4;	DB 6;	Length 128;
	Best Local Similarity	99.2%;	Pred. NO. 9.3e-22;		
	Matches 127;	Conservative	0;	Mismatches 1;	Indels 0;
	Gaps	0;			
Qy	3120	ggagagtgggctcgtcctcatcctagcgcacgcactgattgttacctgtgtgcagaagaa	3179		
Db	1	GGGAGGTGGGCTCTGCTCATCTTAGGCATTCGCACCTGATTGTACCTGTTGCAGAAAGAA	60		

This Page Blank (uspto)